

182 ATGGACGGACCGGTAAATCGA

17 pAsnSerProProgiuProSerProgiuSerleuAtrHisValSera 34
 232 CAATTCCTACCGCTGAGCCATACCGAGAGTCTCCGTATGTAAAGCC 281
 34 rGMeLLeasSerAsnHisTyrThrSerProSerArgThrIleTyrSer 50
 282 GTATGATCAACAGACCACTTACACCTCCTCTCGAACAAATCTACTGCC 331
 51 AsparGhellieProSerArgSerAlaSerLysPhealaleuPheasp11 67
 332 GATAGGTTTCATCCGAGTAGATCTGCTTCGAAATTCGCTTGTGTTGATAT 381
 67 eAsnThrProThrGluArgAspAspSerSerSerAlaTyrThrThrL 84
 382 CAATTCCTCCAGACAGAGACCGCATGATGATTCACGCGCTTATACGACTC 431
 84 euleuArgThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
 432 TTCTGAGAAAGCGCGTGTGTTGACCGGATGTGCGCGTTCACGCGC 481
 101 GluLysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheAr 117
 482 GAAAAAACCGACTCGCCGTCGATGACATTCGCGAATAGAAATATTTTAC 531
 117 qTyrLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMetA 134
 532 GTATTAAGACGAGACAGACAGACATCATGCTCCGCTTCGCGCTTATAG 581
 134 sPAspAspPheValProGlyValAsnHisSerProValLysAlaProArg 150
 582 ATGATGATTTTGTCTCTGTTATCATCTCCGTTAAAGCTTCACAGG 631
 151 LysValProArgSerProTyrLysValLeuAspAlaProAlaLeuGlnAs 167
 632 AAGGTTCTCGATCGCTTATAGGTTTGGATGCACTGCTTGGCAGA 681
 167 pAspPheTyrLeuAsnLeuValAspTyrPheSerHisAsnValLeuAlaV 184
 682 TGATTTTATCTGAATCTGTAGATGTGCTTCAACAAATGTGTTGGCTG 731
 184 aLgLyLeuGlyAsnCyValTyrLeuTyrPAsnAlaCySerSerLysVal 200
 732 TTGGTTTGGTAACTGTCTATCTCTGGAATGCTTGTAGCAGCAAGGTA 781
 201 ThrLysLeuCyAspLeuGlyValAspAspLysValCySerSerValGlyTr 217
 782 ACTAATATTATGATTTGGGGGTGATGATGTGTGTGTGTGTGTG 831
 217 pAlaGlnArgGlyThrHisLeuAlaValAlaGlyThrAsnAsnGlyLysValG 234
 832 GGCTCAACGTGTAATCTGCTGCTGGAATCAACAAATGTAAGTTC 881
 234 InIleTyrAspAlaAlaArgCyLysLysIleArgSerMetGluGlyHis 250
 882 AGATTTGGGATGCACCAAGATGCAAGAAATCAATATGAGAGCGCAT 931
 251 ArgLeuArgValAlaLeuAlaTyrPheSerSerLeuLeuSerSerG1 267
 932 CGGTACTGTGCGGCTTGGCGCTGAGTTCATCTCTTTGCTCTTCGG 981
 267 yGlyArgAspLysAsnIleTyrGlnArgAspIleArgThrGlnGluAsp 284
 982 TGGACGGATTAAGATATTATCAACGAGATATACGACCAACAAGATT 1031
 284 heValSerLysLeuSerGlyHisLysSerGluValCysGlyLeuLysTrp 300
 1032 TTGTAGTAACTGTACAGACCAAAATCAGAGGTTTGGAGCTGAGGTG 1081
 301 SerTyrAspAsnArgGluLeuAlaSerGlyLysAsnAspAsnLysLeuPh 317
 1082 TCATATGATTAACCGTAGTGTGCACTGTGAGAAATGACAACAATTTGTT 1131

317 eValTyrPAsnGlnHisSerThrGlnProValLeuLysTyrCyGlnHisT 334
 1132 TGTTTGGAAATCAACACTCAACCCAGCTGTCTCAAGTACTGTGAGACA 1181
 334 hrAlaValLysAlaIleAlaTyrPheSerProHisLeuHisGlyLeu 350
 1182 CACACACTGTAAAGCTATGTGATGCTTCATCTCATCTTCATGAGACTTCT 1231
 351 AlaSerGlyGlyThrAlaAspArgCysIleArgPheTyrAsnThrTh 367
 1232 GCATCTGGAGAGAACTGCATAGATGATGATTCGTTTGGAAATACAC 1281
 367 rThrAsnSerHisLeuSerCysMetAspThrGlySerGlnValCysAsnL 384
 1282 CACAACTCACACTTACGCTGATGAGACACTGGAATCAGTTTGCATAC 1331
 384 euValTyrPheLysAsnValAsnGlnLeuValSerThrHisGlyTyrSer 400
 1332 TTGCTGTGTCCAAAATGTCAACGAACCTAGTAAGCACATGGATCTCC 1381
 401 GlnAsnGlnIleLeuValTyrPArgTyrProThrMetSerLysLeuAlaTh 417
 1382 CAGAACCAAGATTATGTGTTGAGATACCCCATATGTCAAAAGCTGGCAGC 1431
 417 rLeuThrGlyHisThrTyrArgValLeuTyrLeuAlaIleSerProAspG 434
 1432 TCTTACCGGCATACCTATAGGGTCTCTATCTTGCATCTCCAGATG 1481
 434 LysGlnThrIleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsn 450
 1482 GACAGACTATTGTAACTGAGCGCTGAGATGAACGCTTAGGTTCTGGAAT 1531
 451 ValPheProSerProLysSerGlnAsnThrGluSerGluIleGlyAlaLe 467
 1532 GTTTCCCTCCCTCAATACAGAGATCTGAAGAGTAATGAGACATT 1581
 467 uSerLeuGlyArgThrThrIleArg 475
 1582 ATCTCTTGGAAGAACTACTATACAG 1606
 seq_name: gb_pat:AX008931

seq_documentation_block:
 LOCUS AX008931 2006 bp DNA linear PAT 06-SEP-2000
 DEFINITION Sequence 1 from Patent W09964451.
 ACCESSION AX008931
 VERSION AX008931.1 GI:9996328
 KEYWORDS
 SOURCE
 ORGANISM
 Medicago sativa.
 Medicago sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 1 (bases 1 to 2006)
 Cebolla,A., Kondorosi,E. and Kondorosi,A.
 Plant protein with repeated wd40 motifs, nucleic acid coding for
 said protein, and uses thereof
 Patent: WO 9964451-A 1 16-DEC-1999;
 CEBOLLA ANGEL (FR); KONDOROSI EVA (FR); KONDOROSI ADAM (FR); CENTRE.
 NAT RECH SCIENT (FR)

FEATURES
 source
 location/Qualifiers
 1..2006
 /organism="Medicago sativa"
 /db_xref="taxon:3879"
 182..1609
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC07447.1"
 /db_xref="GI:9996329"
 /translation="MDGNGNPNPPTSTVDRNSPPPEPSPESLRHSRMINSNHTSP
 SRTIYSDFITRSKSKFALFDITPTPEGSDSSATYTLRLALGCPDVGAVTPEK
 TDSPSMTLPNNIRFRTETKQSMHSLSPTNDDDFVGVNHSFPAKRVKRPVTVLV

1082 TCATATGATTAACCGTAGTGTGCACTGTGAGAAATGACAACAATTTGTT 1131

DAPALODDEYFLNLVDMSHNVLAVGLGNCVYLMMNCSKVKTKLCLDGLVDDCVCSGVMA
 ORGTHLAVGTNNKQVIMDAARCKIRSMEGRLRYGLALMSSSLSSGGRDKNTOR
 DIRTHEDPVSCKNGSEVCGKMSYDNRELASGGNDKLFPWNOHSTQPVLYKCEHT
 LAATKATAPSHLGLIASGGGTADRCIRRMNTTNSHLSMDTSGSOVCLMYSKRVNE
 LYSTHYSOQIIVWRYIPMSKATITITGHTYLVLIASIPDQITVTGAGDETLRFNM
 VFPSPKSNQTESIGLSLGRITIR

BASE COUNT 555 a 398 c 441 g 612 t
 ORIGIN

alignment_scores:

Quality: 2555.00 Length: 475
 Ratio: 5.379 Gaps: 0
 Percent similarity: 100.000 Percent identity: 100.000

alignment_block:

US-09-701-572-2 x AX008931

Align seg 1/1 to: AX008931 from: 1 to: 2006

1 MetAspGlyThrGlyAsnArgAsnProProProThrSerThrValArgAs 17
 182 ATGGACGGACCGGTAATCGAATCCACCGACTTCCACCGTCAGAGA 231
 17 PAsnSerProProProGluProSerProGluSerLeuArgHisValSerA 34
 232 CAATTCTCCACCGCTGAGCCATCCAGAGAGTCCCTCCATGTAAGCC 281
 34 rgmErlAsnSerAsnHisThrThrSerProSerArgThrIleTyrSer 50
 282 GATATGATACAGCAACCATTCACCTCCTCCGACCAATCTACCTCC 331
 51 AspArgPheIleProSerArgSerAlaSerLysPheAlaLeuPheAspI 67
 332 GATAGTTCATTCGAGTAGATGCTGCTCGAAATCGCTTGTGTGATAT 381
 67 eAsnThrProThrGluArgAspAspSerSerSerAlaTyrThrThr 84
 382 CAATCTCCGACAGAGACGCGATGATGATGTCAGCGCTTATTCAGACTC 431
 84 euLeuArgThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
 432 TTCTGAGAACGGCTTGTGTGGACCGGATGTGCGCGTCCGGTTACGCC 481
 101 GluIystrAspSerProSerMetThrLeuProAsnArgAsnIlePheAr 117
 482 GAAAAAACCGACTCGCGTCGATGACATTTGCCAATAGAAATATTTTAG 531
 117 gTyrIystrThrGluThrArgInserMetHisSerLeuSerProPheMe 134
 532 GTATTAAGACGGAGACAGACAGTCCATGCTCGCTTCGCCGATTATAG 581
 134 sPAspAspPheValProGlyValAsnHisSerProValLysAlaProArg 150
 582 ATGATGATTTTGTCTCTGTTAATCATCTCCGTTAAGGCTCCTTAGG 631
 151 LysValProArgSerProTyrLysValLeuAspAlaProAlaLeuGlnAs 167
 632 AAGGTTCCTCGATCGCTTATTAAGGTTTGGATGACACTGCTTTCAGA 681
 167 pAspPheTyrLeuAsnLeuValAspTyrSerSerHisAsnValLeuAla 184
 682 TGATTTTATCTGAATCTGTAGATGTGTTCACACAAATGTGTGGCTG 731
 184 aIGlyLeuGlyAsnCyValTyrLeuTyrPAsnAlaCySerSerLysVal 200
 732 TTGGTTTGGTACGTCTCTATCTCTGGAATGCTGTAGACAGCAAGTA 781
 201 ThrLysLeuCyAspLeuGlyValAspAspCyValCySerValGlyTyr 217
 782 ACTAATATATGATTTGGGGTGTGATGATGTGTTGTTCTGTGTGGTTG 831
 217 pAlaGlnArgGlyThrHisLeuAlaValGlyThrAsnAsnIyLysValG 234

882 GCGTCAACCGGTACCTCATCTTCTGTGGAACTAACAATGTAAGTTC 881
 234 InIleTyrPAspAlaIleArgCyValLysLysIleArgSerMetGluGlnHis 250
 882 AGATTGGGATGCAAGATGCAAGATGCAAGATGCAAGATGCAAGATG 931
 251 ArgLeuArgValGlyAlaLeuAlaTyrSerSerSerLeuLeuSerG 267
 932 CGGTACGCTGCGGGGCTTGGCTGAGATCATCTTGTGTCTCTG 981
 267 yGlyArgAspLysAsnIleTyrGlnArgAspIleArgThrGlnLysP 284
 982 TGGACGGATAGAAATATTTATCAAGATATACCAACCAAGAGATT 1031
 284 heValSerLysLeuSerGlyHisLysSerGlyValCyGlyLeuTyr 300
 1032 TTGTTAGTAACTGTCAAGACACAAATCAGAGGTTTGTGACTGAAGTGG 1081
 301 SerTyrAspAsnArgGluLeuAlaSerGlyLysAsnAspAsnLysLeu 317
 1082 TCATATGATTAACCGTGAAGTGGCATCTGGAGAAATGACAAATTTGTT 1131
 317 eValTyrPAsnGlnHisSerThrGlnProValLeuLysTyrCyGlnHis 334
 1132 GTTTGGAAATCAACACTCAACACCGCTGTCTCAAGTACTGTGACACA 1181
 334 hrAlaIleValLysAlaIleAlaTyrSerProHisLysGlyLeuLeu 350
 1182 CAGCAGCTGTTAAAGCTATTGCAATGCTCTCTCATCTTCAATGACTCT 1231
 351 AlaSerGlyGlyLysThrAlaAspArgCysIleArgPheTyrPAsnThr 367
 1232 GCATCTGAGAGAGAACTGCAGATAGATGATGCTTTTGGAAATCAAC 1281
 367 rThrAsnSerHisLeuSerCysMetAspThrGlnLysGlnValCyAsn 384
 1282 CACAAACTCAACCTTAGCTGTATGAGACACTGGAAGTCAGGTTTGCAAATC 1331
 384 euValTyrSerLysAsnValAsnGluLeuValSerThrHisGlyTyrSer 400
 1332 TTGCTGTGTCACAAATATGCAAGAACTAGTAAACACACTGGGTACTCC 1381
 401 GlnAsnGlnIleIleValTyrArgTyrProThrMetSerLysLeuAla 417
 1382 CAGACACAGATTAATTTGTTGAGATACCCCATATGTCTCAAGCTGGCGAC 1431
 417 rLeuThrGlnHisThrTyrArgValLeuTyrLeuAlaIleSerProAsp 434
 1432 TCTTACCGGCATCTATAGGGGTCTATCTTCCATCTTCCAGATG 1481
 434 LysGlnThrIleValThrGlyAlaGlyAspGluThrLeuArgPheTyrPAsn 450
 1482 GACAGACTATTGTAACTGAGCGAGCTGAGATGAAGCGTTAGGTTGCAAT 1531
 451 ValPheProSerProLysSerGlnAsnThrGluSerGluIleGlyAla 467
 1532 GTTTTCCCTTCCCTTAATCACAAGAACTACTGAAAGTGAATTCGAGCAAT 1581
 467 uSerLeuGlyArgThrThrIleArg 475
 1582 ATCTCTTGGAAGAACTACTATACAG 1606

seq_name: gb_pl:AY063875

seq_documentation_block:

LOCUS AY063875 1911 bp mRNA linear PLN 05-DEC-2001
 DEFINITION Arabidopsis thaliana putative Strv1 protein (At4g11920) mRNA.
 ACCESSION AY063875
 VERSION AY063875.1 GI:17380837
 KEYWORDS FLI CDNA.
 SOURCE thale cress.

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1	(bases 1 to 8450)			2	(bases 1 to 8450)		
Medicago truncatula				Kondorosi, A. and Kondorosi, E.			
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Eudicotyledons: core eudicots: Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Trifoliaceae: Medicago.				The mitotic inhibitor ccs52 is required for endoreduplication and polyploidy-dependent cell enlargement in plants			
Unpublished				Submitted (12-MAR-1999) ISV, CNRS, Avenue de la Terrasse, Bat. 23, Gif Sur Yvette 91198, France			
Location/Qualifiers							
1..8450							
/organism="Medicago truncatula"							
/strain="Ghor"							
/db_xref="taxon:3880"							
join<2461..2934,3022..3144,3295..3399,5309..5488,5512..5634,5735..5929,6459..6560,6651..6713,6827..6895,7452..>7505)							
/gene="CCS52"							
/product="CCS52"							
<2461..>7505							
/gene="CCS52"							
join(2461..2934,3022..3144,3295..3399,5309..5488,5512..5634,5735..5929,6459..6560,6651..6713,6827..6895,7452..7505)							
/gene="CCS52"							
/codon_start=1							
/product="WD-repeat cell cycle regulatory protein"							
/protein_id="AAF37386.1"							
/db_xref="GI:7158292"							
/translation="MDGTRNRPPTSTVGDNPPPEPSPESLRHVSRLMINSNHTSPSRIVYDRIPRSKSAKFALFDINTPEGRDSSAYTTLRTALFGDVGVPYPERKTDSPSPTLDNRNIFRYKPTROSMHSISPMDDDPVGINSYKAPKPRSPRYKVALDAVALDDDYLYNLVNDSSHNVLAVGLGNCVYLIMNACSKKVIKLDIGDVCVGVGAORETHLAVGTNNNGKQVIMDAARCKRTIRSEHGRLRVLGALAMSSSLSSGGDKNIYORDIETQDDPVSKLSGHRSYVCGKWSVDNREKLASGGDNKLFVMDHSTQPLVKCEHTAAVKAIAMSPHLGLGLASGGTADRCIRFNMNTTSHLSCHDTSQVCNLIWMSNVMEIVTGHYSNOIIVMRYPMTSKLATLTGHTYRVLYLAISPDGQIVTGAQDETILFVNVPSPSKQMTSEIGVSLGRTTIR"							
BASE COUNT	2503 a	1387 c	1495 g	3065 t			
ORIGIN							
alignment_scores:							
Quality: 1779.00							
Ratio: 3.901							
Percent Similarity: 30.790							
Percent Identity: 30.520							
alignment_block:							
US-09-701-572-2 x AF134835							
Align seg 1/1	to: AF134835	from: 1	to: 8450				
1	Metaspglythgilyasnaargasnproprothrsertthvalatgas	17					
2461	ANGCAGCGGAACCGTAATCGAATCCACACCGACTTCACCGTCGCGAGA	2510					
17	pansnerpropoprogluproserprogluserleuarghisvalsera	34					
2511	TAAITTCACCGCCGCGAGCCATCCGCGAAGTCTCCGCATTAAGCC	2560					
34	rgmellleasnserasnhsrtyrthrsrproserargthrllefyser	50					
2561	GTATGATCAACAGACCATTAACACCTTCTCGAACATCTACTCC	2610					
51	asparqphellleproserarjserlaserlysphealaleupheaspil	67					

```

|||||
2611 GATAGGTGATCCGAGTAGATCTGCTCGAATTCGTTGTTGATAT 2660
67 eAsnThrProThrGluGlyArgAspSerSerSerAlaTyrThrL 84
|||||
2661 CAATACCTCGACGAGACGACGGATGATAGTCCACCGCTTATACGAC 2710
84 euleuArqThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
|||||
2711 TTCTTGAAACGGCGTTGTTGACCGGATGTTCCAGGTCCGGTTACGCCG 2760
101 GluLysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheAr 117
|||||
2761 GAGAAAAACGACATCCGCCGTGATGACATTCGCCAATAGCATATTTTATG 2810
117 gTyrlsThrGluThrArgGlnSerMetHisSerLeuSerProPheMetAl 134
|||||
2811 GATTAAGACGAGACGACAGATCCATGCGCTTCGCGCTTATATG 2860
134 sPAspAspPheValProGlyValAsnHisSerProValLysAlaProArg 150
|||||
2861 ATGATGATTTTGTCTGCTGATTAATCATAGTCCGTTAAGGCTCTAG 2910
151 LysValProArgSerProTyrLys..... 158
2911 AAGGTCCCTCGATCGCTTATAGGATTTCAATAAAAAACTTATCTT 2960
158 ..... 158
2961 TTTTATTTTCTTGATCTTTGTTGAATGTTGTTAGGCTTTTGA 3010
159 .....ValleuAspAlaProAlaLeuGlnAspAspPheTyrLeu 171
|||||
3011 TTTTGATATAGGTTTGTGATGACCTGCTTTCAGATGATTTTATCTG 3060
172 AsnLeuValAspTrpSerSerHisAsnValLeuAlaValGlyLeuGlyAs 188
|||||
3061 AATCGTAGATGTTGCTTCACACATGTTGGCTGTTGGTAA 3110
188 nCysValTyrLeuTrpAsnAlaCysSerSerLys..... 199
|||||
3111 CAGTGTATCTCTGGAATGCTTGTAGCAGCAGAGGTGAGTTTATCAT 3160
199 ..... 199
3161 ATGATTTATGTTGATTTCAAAATTTTCATTTTATGATATGAG 3210
199 ..... 199
3211 CTAGATGGTGTGGCTAGGAAGATTGAACCTGAAGTTGTTTATTA 3260
200 .....ValThrLysLeuCysA 205
|||||
3261 ATTGATTTATTTGATCTTGATATGTTTATGAGAGTACTAATATATG 3310
205 sPLeuGlyValAspAspCysValCysSerValGlyTrpAlaGlnArgGly 221
|||||
3311 AATTGGGGTGTGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3360
222 ThrHisLeuAlaValGlyThrAsnAsnGlyLysVal.Gln..... 234
|||||
3361 ACTCATCTTGGCTGTGGAACTAACAATGTTAAAGTTCCAGTTAAGACTT 3410
234 ..... 234
3411 TGTGCTAATGTTGATCATAAATTAGGTTTCTAATGAGTGATTTTATG 3460
234 ..... 234
3461 TGTATTTGTGCAATTAGAGGTTCTTACGTTTATTTATTTCTGCTGA 3510
234 ..... 234
3511 TTTCTTCTGGGTTTCCCTGCTGTTTATTTTCTGCTCATATATG 3560
234 ..... 234
3561 TTTCTTATTTTCCATTTTGAAGGAATCTTTTCCATTTGGAGTGAC 3610
234 ..... 234
3611 ACGCATTTGATGATTAACAGCAGTGTATTCTTACATTTGTAATAAGT 3660
234 ..... 234
3661 TTATAGCATTTTATCATTTCTAATGTTGATTAAGTAACTGCTATTACTG 3710
234 ..... 234
3711 ATACTGAACCTTATATTTGTTAGACAGCCCTTTTCTGATGACAG 3760
234 ..... 234
3761 GAGCTTCTGTTCTGGGTAGACTCTTACAGATATATACCTTCTATAT 3810
234 ..... 234
3811 ATGTCTAATACATATAGCATTAAGCATTAAGGATATATTAATATTTAAT 3860
234 ..... 234
3861 GAATTTATGTTGATTTATGCAATTTATGATGATACATATATATGTT 3910
234 ..... 234
3911 TTGATAGTTTCTTGTATTAAGTCTGTGATATTTGCCGTAGTACGT 3960
234 ..... 234
3961 CTGTGCTGTTGCTGTGATTTGTTGATGATGTCGGTGTGTGCTGTT 4010
234 ..... 234
4011 GTGATTAATGCTATATATTCATTTATGTCATGTTGTCATTTGCTATAT 4060
234 ..... 234
4061 TTAATGCCGTGCTGCTGCTGTTGTGATATGCTATATATTCATTCGTG 4110
234 ..... 234
4111 CCGTGTGTCATTTGCTATATTAATGCGGTGCTGTTGTCATTTAT 4160
234 ..... 234
4161 GTCTAATTCATTTCTGTCGTGTTGTGATTTGTCATTTATATATGCG 4210
234 ..... 234
4211 TGTGCTGTGTATATATGCTATATTAATGTCGTGCTGTTGTTAT 4260
234 ..... 234
4261 ATGCTATATTAATTTTATATGTAATCTGACTTCACTTACTCTTCT 4310
234 ..... 234
4311 TGATCATCAAAAGTTCTGTTACTATAAAGTTGGTCAATGTGACCGTAT 4360
234 ..... 234
4361 ACCAGCTAACACATCTGCAATTAATTTATATATTCACATGATATTA 4410
234 ..... 234
4411 GTTCAATTATGTTTGAATTCGAGCTTGTGCTTGTATAGGCTTCAC 4460

```

234 234
 4461 CATTAACCCATGAGCGCATGCTCTACATATGACCAGTGGCACTCA 4510
 234 234
 4511 TGTTCGATGTCMAACCCATTGGGTMAAAAAATTAGTAGCAATACTT 4560
 234 234
 4561 CGATACATATGATGAGTACTAGAACCAATCTTAACACCAATATGATTA 4610
 234 234
 4611 GAGGGGAGAGTCAAGTCTTTATGATATTCACACCTGGAACACGCTG 4660
 234 234
 4661 ATGCATGATCATGCTCTGTTAGCTAATCATTTAATGAATAGTTAAT 4710
 234 234
 4711 AATTAATGTTCTGTACTGTTTCTGCGAGAAAAAATATGACAGC 4760
 234 234
 4761 GTTCTGTATTAATCTGTAATATCTGCAGTCAACCTGACCTGAGG 4810
 234 234
 4811 TTCTACAGATCTAGAAATGACGCTTTAATAAAACGGGCTGAAGCA 4860
 234 234
 4861 ACTTAGATCTGGTTTGGCAACTGTAAGTCTGATAGCTATCTACTG 4910
 234 234
 4911 AGATAGTCTATACCTTTAGGAGGCTCTATATTTGACTTTTAACGAGA 4960
 234 234
 4961 ACACCTAATTTGCTTGTGATGTAATCTAATCAATCAATCAATAC 5010
 234 234
 5011 ACATTCCTCCCAAGAGTGTATATAGATTAACCTGTTTCTCTACCA 5060
 234 234
 5061 TAGTATATAGAAAGTGCATCTCCATCTAATATTTTCCTTTGAGAC 5110
 234 234
 5111 TGCCTCCCATGAAAGTGTATATTAAGTGAATCTGCTTAACCTGCTC 5160
 234 234
 5161 TGGTATCAAAATATGAATATCTAATTTCCCATGATCCCTGGAAGT 5210
 234 234
 5211 TCGGAAGTCTACTATCTTTTTTTTAGGAAACTTAATGCTCTCTGTTT 5260
 235 235
 5261 CTATGCTGCAAAATTAACATGCTGTTGTTTTTACTATATCATGTAGAT 5310
 235 235
 5311 TTGGGATGACGACAGTGCAGAAAGTAAATGATCAATGAGGCGCATCGT 5360

252 euArgValGlyAlaLeuAlaIleuAlaIleuSerSerSerLeuSerSerGlyGly 268
 5361 TAGGTGCTGGTCTGCTGGCTGGAGTTCATCTTTCTTCTGCTGGTGA 5410
 269 ArgAspLysAsnIleTyrGlnArgAspIleArgPheGlnLysPheVal 285
 5411 CGGATAGAGATATTTATCAACGAGATATACGACACAGAAAGATTGTT 5460
 285 IserLysLeuSerGlyHisLysSer 293
 5461 TAGTAACCTGTCAGACACAATCAAGAGGTAGATTAGTCTATTTCTT 5510
 293 293
 5511 AGCTTCATCACTTATACATGCTATTCATCAATGATTTGATTGATTTT 5560
 294 Glu. ValCysGlyLeuLysTrpSerTyrAspAsnArgLeu 307
 5561 ATTTTTCGAAGTTTGTGAGCTGAAGTGTCAATGATATACCGTGTG 5610
 308 AlaSerGlyGlyAsnAspAsn.Lys 315
 5611 GCATCTGAGAAATGACAAACAAGATGCAATCAACAACTTAATGATT 5660
 315 315
 5661 AUCTGAAGTATCCACTGCATTAATTTCTTTCTCTCTGACAGGAA 5710
 316 LeuPheValTrpAsnGlnHisSerTh 324
 5711 CTCACATCAATTCATATTGTTGAGCTGTTGTTGTTGGAATCAACACTCAAC 5760
 324 rGlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleA 341
 5761 CCAGCTGTCTCCCAAGTACTGAGCAGACAGCGCTGTTAAAGCTATTG 5810
 341 IatrpSerProHisLeuHisGlyLeuLeuAlaSerGlyGlyThrAla 357
 5811 CATGTCCTCTCATCTTCATGAGACTCTTGCATCTGAGAGAGAACTGCA 5860
 358 AspArgCysIleArgPheTrpAsnThrThrAsnSerHisLeuSerC 374
 5911 TATGACACTGGAAGTCAAGTAAGTACTTCTTAACCTGTTATATTTT 5960
 380 380
 5961 CATACAAATGCACCTGTATTCCAATTTGCTGTATTTGGCGCGCTGGT 6010
 380 380
 6011 TGTGGCCTATATAGTTGTTGCTAGTCTAGAAAACCTATATTTAAT 6060
 380 380
 6061 TTGTTGGAAGGGAAGGGGAATGCTACTTCCCAACACACTTCCCTCT 6110
 380 380
 6111 CAGTTTTCCTCACTTACACAGCATTTTATGAGTATGAATATCAT 6160
 380 380
 6161 TCCTTACTTATCATATATAGTATGACTCAGACATCACTGGGAAATGGCTC 6210
 380 380
 6211 CAGCTAGATTCTTACTATTAAGGTTGTAATAAATCTTAATA 6260
 380 380

6261 AAACGAAAGCAGATCCAGTCTAGTCTAGCTGGTACCATTCATATCACTGCAAAA 6310
380 380
6311 GTTAGATATTGAAAGTCTTTTTCAGTTCATTAAGTTTCAGATATATAGT 6360
380 380
6361 ATTGCGCTTACATATTGAGAGGAAGATAATAGTCTTCTGCTCTTT 6410
381Va 381
6411 CTCCTTCTTAACACAAATCAATGATTTTATCTATATATTTTTCAGAGT 6460
381 lcyasnlenuvaltyrserlysasnlvalasnlenuvalserthrhisg 398
6461 TTGCATCTTCTGCTGCCAAAATGTCATGAAGTCTAGTAGGACACACTG 6510
398 ltyrserglnasnlnllelvaltyrparqtyrprothmetserys 414
6511 GCTACTCCGAGAACCCAGATATGTTGGAGATACCCACATATGTCAAG 6560
414 414
6561 GTACAAAGTGTCTTCTGTGATTAAGTATGCTCTTTGTTATTCACCTAAG 6610
415leu1athrl 418
6611 TTAAACAAGTTTTCACATATTTTTCCTTTCAGTTCAGCTGGCAGCT 6660
418 euthrgllyhsthtryargvalleutyrleualalieserproaspoly 434
6661 TTACCGGCACACTTATAGGGTCTCTATCTTGCCATCTCTCCGAGTGA 6710
435 Gln..... 435
6711 CAGGTATTTAGTCTTTCACATACTCCTCATTTCTAGATTACTAGATATGT 6760
435 435
6761 AATGACAAAGAAATATTTTATATCATCTATGCTATATCTGATGTG 6810
436Thr..llevalthrlyalaglyaspgluthrleua 447
6811 ATGAGTACTATTTCAGACTATTTGTACTGAGCTGGAGATGAACCGCTTA 6860
447 rphetrpaspvalphaproserprolysergln 458
6861 GGTTCGGAATGTTTTCCTTCCCTTAATCAACAG 6895
seq_name: gb_pl:ATF7H19
seq_documentation_block:
LOCUS ATF7H19 139316 bp DNA linear PLN 27-AUG-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F7H19 (ESSA
project).
ACCESSION AT031018
VERSION AT031018
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 139316)
Beyan,M., Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,M.,
Bancroft,I., Mewes,H.W., Mayer,K.F.X., Lemcke,K. and Schneijler,C.
Unpublished
2 (bases 1 to 139316)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (21-JUN-1999) MIPS, at the Max-Planck-Institut fuer

COMMENT

Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/ch1al/Location/Qualifiers>

FEATURES

Location

source
 location/Qualifiers
 1. 139316
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 complement(915. .1445)
 /gene="F7H19.10"
 /number=1
 complement(915. .1445)
 /gene="F7H19.10"
 /note="similarity to PVPR3 protein, Phaseolus vulgaris,
 PID:g169363
 contains EST gb:AA597458, H36610"
 /codon_start=1
 /product="putative protein"
 /protein_id="CA19798.1"
 /db_xref="GI:3292808"
 /db_xref="SPTREMBL:O49663"
 /translation="MGSEQNDSTSTFQSOASEPKLCYKGGFSPSNMDLCKCYNG
 IGEENOTAVAKAAVSEKSPSPRLFLAEPPAVVREPREKAAVYVNSPESSAV
 PEANEPSRAPRTNRCICCKKKYINGFKKCGSTGCEHRIPEPHDCSFDKEVGE
 IAKANVYVAADKIQR"
 complement(915. .1445)
 /gene="F7H19.10"
 915. .1445
 /gene="F7H19.10"
 4680. .6720
 /gene="F7H19.20"
 complement(4680. .4751)
 /gene="F7H19.20"
 /number=1
 complement(join(4680. .4751,4828. .4910,5004. .5257,
 5355. .5607,5685. .5858,5956. .6100,6319. .6720))
 /gene="F7H19.20"
 complement(join(4680. .4751,4828. .4910,5004. .5257,
 5355. .5607,5685. .5858,5956. .6100,6319. .6720))
 /gene="F7H19.20"
 /note="similarity to Na(+) dependent transporter, Aquifex
 aeolicus, PIR2:E70482"
 /codon_start=1
 /product="putative protein"
 /protein_id="CA19799.1"
 /db_xref="GI:3292809"
 /db_xref="SPTREMBL:O49665"
 /translation="MSVITTPILFTLHLKSTRLRLPRAVYSORLQVPPNPISNTLS
 SPLRIDIPISGIFILSSITLRLRLILGTCKTYTFEFSGLQPTIYYGGRNLMRYA
 SDNSSEKDLDEADFEKCAAGSFLFHSMLCWVERPSIVDRKKKANSILIPHYLAST
 ILALYIPGSFTWFSRPELEQGLMEAVGNSNKEQLEAFRRPAKILALGYGVGLV
 KEVLQIFGLAVALSFOLPTPIGAGIMLVGSCVSAQLSNATFRLDPALALSGIVTS
 LSTAAVLAVPEMLSLILGKRLPDVKGIMSSILQVVIAPVIAAGLILNKLKFPVNSAI
 RPELFLIVLDDACVAGPALAINSVMSPEGATILLIYTMFHLNSAFIAGYPLTGSV
 RNAPPAKAKMORTLSTETGMSSLLALALATKFDODPLVGIPIAISTVWMLKGTLLV
 IWSKESNTF"
 complement(4752. .4827)
 /gene="F7H19.20"
 /number=1
 complement(4828. .4910)
 /gene="F7H19.20"
 /number=2
 complement(4911. .5003)
 /gene="F7H19.20"
 /number=2
 complement(5004. .5257)
 /gene="F7H19.20"


```

118 rLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMetAspA 135
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26161 TAAGACGGAGACTCATCGCTGCTTGAATTCCTTCTCCTTTGGCTGG 26210
135 sPaSpheValProGlyValAsnHisSer...ProValLysAlaProArg 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26211 ATGATATTCCTCGTGGTTCCTCATAGTGCCTGTAAAGCTCCACG 26260
151 LysValProArgSerProThrLysVal..... 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26261 AAAGTCCCGGATCGCGTAAAGTGACGATTTTGATTCGGGGTGA 26310
159 ..... 159
26311 TACATTTAGATTTTGGTGTTCATCATGAATCTGATCAGTCCTT 26360
159 ..... 159
26361 GTGTTTTTAAGTGGGTGTCGATTCGATTCGATTCGATTCGATTC 26410
159 ..... 159
26411 TTCGATGGGTTTTTGTGTAATGCTGATGATTTGGGGGTGGAGCTTA 26460
159 ..... 159
26461 GATTCCTGATTCGCTGACCTTATGATCTTGGTTTCGATTAATGATGAA 26510
159 ..... 159
26511 CAATTTGATCTTGTGATGATTTGGTCTGAGGGTCTAGAAATTTGAG 26560
159 ..... 159
26561 TCGAGATTATGATCTTAGATGAACATATTATTGAAGATGCTTGTG 26610
159 ..... 159
26611 TTTTACTGTCCTAGTGTGTGTAATTTAGAAAAGTTAATCTGTT 26660
159 ..... 159
26661 ACTGGTTGGGTAATTTGTTGGGCAATGTCCTACATCTGATTAACA 26710
159 ..... 159
26711 GAGGTTTTCAGTACTTACTAGTGTAGAGTGAATCTAGATGATGAAGA 26760
159 ..... 159
26761 TGTGTTGTTTACGTTTTTTCGATAGTTTTTGTGTGTGTTGAATG 26810
159 ..... 159
26811 CAAACCTGATGTTGTTATGATTTGGATTTTGTATTAGCAATG 26860
160 ..... 160
26861 TTGCTACAGATGATTAACAGAGCTTCTTCTGTTTATGATGATG 26910
162 IAProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTrpSer 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26911 CACCGGCTTCGACGATGATTTTATCTGAATCTTGATTCGATTCG 26960
179 HisAsnValLeuAlaValAlaGlyLeuGlnLysAsnValTyrLeuTrpAsnAl 195
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26961 CAAAATGTTTACGACGTGGAGCTAGGAACGTGTATTTATGGAATGC 27010
195 acySerSerLys..... 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27011 TTGTAGCGCAAGGTGAGTTTTTTTGTGCACTTGCAAAATTTCTAATAAT 27060

```

```

199 ..... 199
27061 CACAGTACTTCAAAATTTTGTGTTGTTGATTCCTCTTCAAAATC 27110
200 .....ValThrLysLeuCysAspLeuGlyValAspPcysValCysSe 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27111 CAACGAGTACTAAGTTATGATCTCGAGCTCGAGATATGTTGCTC 27160
214 rValGlyTrpAlaGlnArgGlyThrHisLeuAlaValAlaGlyTrpAsnArg 231
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27161 AGTGGTTGGCGGTACGTCGAGACTCATCTGCTGTGGAACTAGTACCG 27210
231 LysValAlaGln..... 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27211 GGAAGTTACAGTTAGATCATCTTACTCTTGTGTTTGTCTGTACTA 27260
235 ..... 235
27261 AGAATTTACGTTACTTTCACACTTTCGATTTAGAAAACATATTTAAA 27310
235 ..... 235
27311 GTAACTATTTATAGTACAACAAGATTAATTAAGCTTATCGTGGAC 27360
235 ..... 235
27361 TGTAGGATTCCTGGCATATCTTTACTGCTATGATCTTGATCATTA 27410
235 ..... 235
27411 TATGATTAATTAACAAGATTTTCTTATGAGAAGATATCTGATTA 27460
235 ..... 235
27461 AGCATTAACAAGGCTGTAAATTTTGAAGTCCCTGAATGAACTA 27510
235 ..... 235
27511 AATCCTTTGTGTTAAGATGGAAGTATCTGCATATAAACCTTGGCGGA 27560
235 ..... 235
27561 TAAATTTTCAATATATCTGAAGAATTTATGATGTTCTTCTATGCAAT 27610
236 ..... 236
27611 TCACCTTAGTTGTTGTTTGTAGATATGAGATGCGTACGCTCAAGAGA 27660
244 IleArgSerMetGlnGlyHisArgLeuArgValAlaLeuAlaTrpSe 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27661 ACAAGAACATGGAAGGTCAATCGTCTAAGATTGAGCCCTGCAATGGG 27710
260 rSerSerLeuLeuSerSerGlyArgAspLysAsnIleTyrGlnArg 277
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27711 TTCACTCGTCTGTCATCTGTCAGACAGACAGACAGATTCCTCAGAGAG 27760
277 spLeuArgThrGlnGlnAspPheValSerLysLeuSerGlyHisLysSer 293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27761 ACATTAAGCTCAAGAGATCATGTCAGTAATTTGGAGGTCAATAATCT 27810
294 GluValCysGlyLeuLysTrpSerTyrAspAsnArgGlnLeuAlaSerG 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27811 GAGGTATGCGGACCTCAAGTGTCTTATGACAACAAGAGCTACATCTGG 27860
310 yGlyAsnAspAsnLys..... 315
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27861 TGGAAACGACAAATAGGTATGATGATCAATGATTTGGGTTACATTA 27910
315 ..... 315
27911 TCTTCTTTGTAACCCCAACATAGCTAAAGCTTATTAATTTTGCAG 27960
316 LeuPheValTrpAsnGlnHisSerThrGlnProValLeuLysTyrCysG 332

```

```

|||||
27961 CTTTGTGATGGACACATTCAACACACCGGTTTGAATATAGTCA 28010
332 UHSTHRAAAlaValysAlaIleAlaTrpSerProHISLeuHISGLYL 349
|||||
28011 ACACACTGACGCTTAAAGCCATTGCTGCTCCTCATGTTCAATGGCG 28060
349 euleuAlaSerGlyGlyThrAlaAspArgCysIleArgPheTrpAsn 365
|||||
28061 TTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 28110
366 ThrThrAsnSerHisSerCysMetAspThrGlySer..... 379
|||||
28111 ACACACGCAATATCAATTAAAGTTCATAGTACTGACAGTCAAA 28160
379 ..... 379
|||||
28161 TTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATG 28210
380 ..... GluValCysAsnLeuValTrp 386
|||||
28211 TTAATATTCATCTTTGCTTATATATATATATATATATATATATAT 28260
387 SerLysAsnValAsnGluLeuValSerThrHisGlyTrpSerGluAsnG 403
|||||
28261 TCTAAGACGTAACGAGCTTGTAGACACACGAGATCTCAGAACCA 28310
403 nIleIleValTrpArgTrpProThrMetSerLys..... 414
|||||
28311 AATCATTTGCTGGAATATACCAACATGTCGAAATATCTCTCGA 28360
414 ..... 414
|||||
28361 CCTATATATCTTCCATCATGATGATGATGATGATGATGATGATGAT 28410
415 ..... LeuAlaThrLeuThrGlyHisThrTrpArgValIle 426
|||||
28411 TTTTCGTTGTCACATGCTCTTAAACCGTCTCACATACCGAGTCTT 28460
426 uTrpLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaG 443
|||||
28461 ATACCTGGGCTTTCACCCGATGACAGACGATGTAACAGAGAGAGAG 28510
443 spGluThrLeuArgPheTrpAsnValPheProSerProLysSer.... 457
|||||
28511 ATGAACCTTAAGGTTCTGGAATTTTCCCTCCCAAAATCTCAGCTA 28560
457 ..... 457
|||||
28561 AAAACCACTACTCTGCTGTAATTACTTAATGATCACAATAATGCT 28610
458 ..... GlnAsnThrG 461
|||||
28611 ATTCTCAGAAAGCTGATTCGATGTTTGGCTGTTGTTGCAACACAG 28660
461 InSerGluIleGlyAlaLeuSerLeuGlyArgThrThrIleArg 475
|||||
28661 ATATGAAATCGGTCGCTTTCTTGGTAGAACAAATTCGG 28704

```

seq_name: gb.pl:ATCHRIV58

seq_documentation_block:

LOCUS ATCHRIV58 193550 bp DNA linear PLN 16-MAR-2000
 DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 58.
 ACCESSION AL161558
 VERSION AL161558.2 GI:7269123
 KEYWORDS

SOURCE

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 29088 to 168403)

AUTHORS Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
 JOURNAL Unpublished
 REFERENCE 2 (bases 154197 to 193550)
 AUTHORS Massen, O., Clabault, G., Quigley, F., Mache, R., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 38442)
 AUTHORS Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
 JOURNAL Unpublished
 REFERENCE 4 (bases 1 to 193550)
 AUTHORS EU Arabidopsis sequencing, project.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, project coordinator: Mike Beyer, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UD Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk
 COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with ATCHRIV57 at the 5' end and an overlap with ATCHRIV59 at the 3' end.
 FEATURES
 source Location/Qualifiers
 1..193550
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 6009..7569
 /gene="AT4g22770"
 complement(6009..6398)
 /gene="AT4g22770"
 /number=1
 complement(join(6009..6398,6575..6706,6791..6858,7155..7569))
 /gene="AT4g22770"
 complement(join(6009..6398,6575..6706,6791..6858,7155..7569))
 /note="similarity to DNA-binding protein PDI, Pisum sativum, PARCH: E27518 contains EST gb:N37194"
 /codon_start=1
 /product="putative DNA binding protein"
 /protein_id="CAW79232.1"
 /db_xref="GI:7269124"
 /translation="METHGEVYKTTGSDGVTYVRSNAPDFFHMAPRSETSNTPNS VAPPPPPONSFPSSAMDGSSGPIKRRGRPKYGHDAVAVTLSPNPISAPPT SHVIDFTSTSEKRMKPAETPPSSFIKPKYOVLEKSGSPSSAANTPIIIVNAG EDVKRIISFOGSLACVLCANGVSVTLRQDSSGGLTVEGREILISGTFM PSDSGTFRSGMGVSLASPDGVRVGGVAGLLVAAATPIOVVGTFLGCTNOGQRP KPHNNPMSPLMPSSNVADHRTIRPMTSLPSTWTFPSDSRKHSHDENITTL "
 complement(6399..6574)
 /gene="AT4g22770"
 /number=1
 complement(6575..6706)
 /gene="AT4g22770"
 /number=2
 complement(6707..6790)
 /gene="AT4g22770"
 /number=2
 complement(6791..6858)
 /gene="AT4g22770"
 /number=3
 complement(6859..7154)
 /gene="AT4g22770"
 /number=3
 complement(7155..7569)
 /gene="AT4g22770"
 exon
 Intron
 Intron
 exon

```

exon      /number=4
           complement(10826..11212)
           /gene="At4g22780"
gene       /number=1
           /gene="At4g22780"
           complement(join(10826..11212,11306..11977,12062..12129,
           12220..12334,12424..12496,13025..13086))
           /gene="At4g22780"
           complement(join(10826..11212,11306..11977,12062..12129,
           12220..12334,12424..12496,13025..13086))
           /gene="At4g22780"
           /note="similarity to translation factor EF-1 alpha
           gamma family, Arabidopsis thaliana, PATCHX:G1532164
           contains EST gb:217780"
CDS        /codon_start=1
           /product="translation factor EF-1 alpha-like protein"
           /protein_id="CAB79233.1"
           /db_xref="GI:7269125"
           /translation="MELSDCSNEYKLVYRMNMPRVVIDNGVCPNSTVVKVFTLTQPS
           SIDASRSPGILLESVOLFLDMNLWIKKAVISSDGKMMNDVFNVDLNGDKLIDENLIR
           YLEKSTESHVCKTEGTYGTLTALFETGDRVGLSEVFAVLADLDCDVEAKWAMHNG
           RIASMTYKDGNSGPIIDSDSPVQVREBQNLNLKADGYNDRPYCSYSGNTHME
           RLIHQEMENDRDYERKEKSPIVSQMLPRGYSVNLQCKDRMLLFDVYCTLID
           MAYIVFHAIRTVGETAFLEFYVRHSDGHPVSEPERQLDQLOAAIERRTKVGVL
           ELCTADRPGLAEVTRILRENLINIAKEISYKDSIARVVFYTDANGNLIDPEIKS
           IREKIGIDLSVKEPPIISCREAVEKEQHIEFDHQGRYGGTVALVSGSLVMRLYH
           LGLIKSYF"
           complement(11213..11305)
           /gene="At4g22780"
           /number=1
           complement(11306..11977)
           /gene="At4g22780"
           /number=2
           complement(11978..12061)
           /gene="At4g22780"
           /number=2
           complement(12062..12129)
           /gene="At4g22780"
           /number=3
           complement(12130..12219)
           /gene="At4g22780"
           /number=3
           complement(12220..12334)
           /gene="At4g22780"
           /number=4
           complement(12335..12423)
           /gene="At4g22780"
           /number=4
           complement(12424..12496)
           /gene="At4g22780"
           /number=5
           complement(12497..13024)
           /gene="At4g22780"
           /number=5
           complement(13025..13086)
           /gene="At4g22780"
           /number=6
           complement(13087..13175)
           /gene="At4g22780"
           /number=1
           17283..18758
           /gene="At4g22790"
           complement(17283..18758)
           /gene="At4g22790"
           complement(17283..18758)
           /note="similarity to predicted protein, Arabidopsis
           thaliana
           contains Prokaryotic membrane lipoprotein lipid attachment
           site AA337-347"
           /codon_start=1
           /product="predicted protein"

```

```

protein_id="CAB79233.1"
/db_xref="GI:7269126"
/translation="MSETSRSESLDEPVESEGLCKTLMOISYHELKLMQIRIGLPLVYM
NLMLFQKMTTSVFLGROGELINLAGEISGFSPANVGFSLVYISAMBEICGAAGA
KNFKLHKTEFMVAILLLISVPSIFMLVHKILGFGOREDISFLAKYLLYLPE
LPIISFCPLKAVIISGCVTLPIMTTAAATSHIPINIVLSKARIEGAVAMAVITD
FTVILITGVYVERKERRKKGKGLNNSADNLTILKSGPCCLTYCLEMCKEI
LVLTGLPMPVQAVSLIIVFENDVLTAVMLSLGCVATRVSNELGANNRGAIRRA
AYTTLIVGIIISGIGALVMTAFRGFWMSLVTTHDQILNCGVKKMLIMAVIEVNPPL
MVCEIIVGTAKPSIGMAYANLSGFYLAIDPLGATVLAFAKQGLQFLIGLFWISLCL
SLIIFARIIDWEKAKQIILNCFDEQTSOGSGDSHS"
complement(20135..20293)
/gene="At4g22800"
/number=1
complement(join(20135..20293,21204..21272,25157..25201))
/gene="At4g22800"
20135..25201
/gene="At4g22800"
complement(join(20135..20293,21204..21272,25157..25201))
/gene="At4g22800"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB79235.1"
/db_xref="GI:7269127"
/translation="MSLTIKPRPRRERKAFGYSGRYFRVFRPSLSEPRVYISRI
DEKLDLSLDVTPDLSEFLLSYFISLIPILKVDKPKCYLLFSS"
complement(20294..21203)
/gene="At4g22800"
/number=1
complement(21204..21272)
/gene="At4g22800"
/number=2
complement(21273..25156)
/gene="At4g22800"
/number=2
complement(25157..25201)
/gene="At4g22800"
/number=3
25157..25201
/gene="At4g22810"
/note="similarity to DNA-binding protein PDL, Pisum
sativum, PATCHX:E275188
contains EST gb:N96206"

alignment_scores:
  Quality: 1562.00      Length: 966
  Ratio: 3.633          Gaps: 17
  Percent Similarity: 44.513  Percent Identity: 37.681

alignment_block:
  US-09-701-572-2 x ATCHRIIV58 ..
  Align seg 1/1 to: ATCHRIIV58 from: 1 to: 193550
  12 ThnSerThValArGAspAsnSerPro.....ProProGluProse 25
  :::::||||:||||:||||:||||:||||:||||:||||:||||:
  54922 TCCTCATCTATGAGAAACCTATCGCGGAGAGATACCTCGGTGTTTC 54971
  25 rProGluSerLeuArGHisValSerArGMetIleAsnSerAsnHisTyrT 42
  ||||| ||||| :::::::::::||||:||||:||||:||||:
  54972 ACTTGAAGTCA.....CGAATCAATCGATTATCAATGCTATCAATCC 55015
  42 hSerPro.....SerArGThrIleTyrSerAspArgp 53
  ||||| ||||| :::::::::::||||:||||:||||:||||:
  55016 AATCACACATCACCATCATCATCAAGGCTATATACCTGATGATTT 55065
  54 IleProSerArGSerAlaSerIysPheAlaLeuPheAspIleasn...Th 69
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  55066 ATCCCAAGTAGATCCGATCCATTCGCTTTTGATCTATTCCTCTTC 55115
  69 rProThr...GluGlyArGAspAspSerSerSerAlaTyrThrLeu 85
  ::|||::: |||||:::||||:||||:||||:||||:||||:
  55116 TCCTAGTAAGATGATGAGAGATGAGCTGCTCTTACGCTACTCTGT 55165

```

85 euargThrAlaLeuPheGlyProAspValAlaGlyProValThrProGlu 101
 |||||.....
 55166 TGGGTGGGGATGTTGGCTGAG.....ACGCCGGAG 55200
 102 LysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArgTy 118
 |||||.....
 55201 AAGAGAGATATTACTGGGTTTCT...TCTCCAGAAATATTTTAGGTT 55247
 118 rLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMetAspA 135
 :|||.....
 55248 TAAGACGAGACATCATCGCTTGAATCTCTTCCTTTGGGTGG 55297
 135 spAspPheValProGlyValAsnHisSer...ProValLysAlaProArg 150
 |||||.....
 55298 ATGATGATTTCTCCGATGTTCTCATAGTGGTCTCTTAAAGCTCCAGG 55347
 151 LysValProArgSerProTyLysVal..... 159
 |||||.....
 55348 AAAGTCCGGGATCGCGGTAAAGTGAGCATTTTGATTCGGGGTGA 55397
 159 159
 55398 TACATTTGATTGTTTGGTGTTCATCATGAACTGATCATGCTT 55447
 159 159
 55448 GTGTTTTTAAGTGGTGTGTGCATGTGTGAATTGATGTTTTTTC 55497
 159 159
 55498 TTCGATGGGTTTTTGTAAATGCTGTAGATTGGGGTTGGGAGCTTA 55547
 159 159
 55548 GATTCCTGATCTGCTGACTTATGATCTTGGTTGCATATGCAAGAA 55597
 159 159
 55598 CAATTTGATCTTTGATGATGTTTGGTCTCTGAGGCTAGAAATTGAG 55647
 159 159
 55648 TCTGAGGTAAATGATCTTATGATGAACATTAATGAAGATGCTGTTG 55697
 159 159
 55698 TTTTACTGTTCTTAGTGTGTGTATTTAGAAAGTTAAATCTTGT 55747
 159 159
 55748 ACTGGTTTGGGTAATTTGTTGTGGCAATGTTCTACATCTGATACA 55797
 159 159
 55798 GAGGTTTGAAGTCTTACTACTAGTAGAGTATCTAGATGATGAAGAA 55847
 159 159
 55848 TGTGTTGTTTACTGTTTTCATAGTTTCTAGTGTGTTGTAAGTGC 55897
 159 159
 55898 CAAAGCCGATGTTGTTATGATTTTGATTTTGTATATAGCAATG 55947
 160Leu,AspA 162
 55948 TTGCTACAGATGATTAACAGAGCTTCTCTGTTTGAAGTATTTGATG 55997
 162 laProAlaLeuGlnAspAspPheTyLeuAsnLeuValAspTrpSerSer 178
 |||||.....
 55998 CACGGGCTTCAAGATGATTTTATCTGAATCTGTGATGTGCTGCA 56047

179 HisAsnValLeuAlaValGlyLeuGlyAsnCysValTyLeuTrpAsnAl 195
 :|||.....
 56048 CAAATGTTCTAGCAGTGGAGTACAGAACTGTGTATTTATGGAATGC 56097
 195 acYserSerIys..... 199
 |||||.....
 56098 TTGTAGCAGCAAGGTGAGTGTGTTTGTGCTGCAAAATTTCTAATAAT 56147
 199 199
 56148 CACAGTACTTCAAAATTTGTTTGGTGTGATTTCTCTCTCAAAATC 56197
 200ValThrLysLeuCysAspLeuGlyAlaAspAspCysValCysE 214
 |||||.....
 56198 CAACAGGTTACTAAGTATGATGATCTCGAGCTGAGATAGTGTGCTC 56247
 214 rValGlyTrpAlaGlnArgGlyThrHisLeuAlaValGlyThrAsnArg 231
 |||||.....
 56248 ACTGGTGGGGCTTACGCGAACATCATCTGGCTGTGGAACCTAGTCCG 56297
 231 LysValGln,Ile..... 235
 |||||.....
 56298 GGAAGTTAGGTTAGATCTCATTTACTCTTGTGTTTGTCTGTACTA 56347
 235 235
 56348 AGAATTTACGTTACTTTCGCACTTCTGATTTGAAAAACAAATTTTAA 56397
 235 235
 56398 GTAAACTATTTTACTTACACATGATTAAGATTAGACTTATCTGTGAC 56447
 235 235
 56448 TGTAGGATTTGCCATATCTTTACTGTATGATGATCATTAAGTGG 56497
 235 235
 56498 TATGATTTACAAAGATTTTCTTATCGAAGAAATGCTCATTTAA 56547
 235 235
 56548 AGCATTAACAAGGCTTGAAGTATTTGGAAGTCCCTGAATGAAGCA 56597
 235 235
 56598 AATCCTTGTTGTTAAGATGAAGAAGTATCTGATATAAACTTGCGGA 56647
 235 235
 56648 TAAGATTTCAATATATCTGAAGAAATTTATGATGTTCTTATGCAATAT 56697
 236TrpAspAlaAlaArgCysLysLys 243
 |||||.....
 56698 TCACCTTACTTGTGTTTGTGATATGAGATGCGTCAGCGCAAGAGA 56747
 244 IleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSe 260
 |||||.....
 56748 ACAAGAACAAATGAGGTCATGCTCAAGAGTTGGAGCCCTGCAATGGGG 56797
 260 rSerSerLeuLeuSerSerGlyValArgAspLysAsnIleTyArgIleArgA 277
 :|||.....
 56798 TTTCATCGGTTCTGTCATCTGTAGCAGACAAAGATATTTCTTCAGAGAG 56847
 277 spIleArgThrGlnGluAspPheValSerLysLeuSerGlyHisLysSer 293
 |||||.....
 56848 ACATTAAGTGTCAAGAAGATCATGTCAATAATTTGGCAGCATATAACT 56897
 294 GluValCysGlyLeuLysTrpSerTyAspAsnArgGluLeuAlaSerG 310
 |||||.....
 56898 GAGGTATCGGAGCTCAAGTGTCTTATGACAAACAGAGACTAGCATCTGG 56947
 310 yGlyAsnAspAsnLys..... 315


```

47 ThrIleTyrSerAspIleProSerArgSerAlaSerIleVal 63
76749 CCAATTACTCAGATAGGTTTATTCAGATGATCTGTTCCAAATTCCG 76700
63 AleuPheAspIle.....AsnThrProThr.....GluGlyArgAsp 76
76699 GCTTTTCATTTGGCATCTCTTCGCCGAAATAGAAAGATGGAGAAAG 76650
76 spSerSerAlaTyrThrIleuValArgThrAlaLeuPheGlyPro 92
76649 ATGGGGCTGGTCTTATCCAGCTCTTTGAAACGGCGCTTTT..... 76606
93 AspValAlaGlyProValThrProGluLysThrAspSerProSerMet 109
76605 .....GGTCCGGTACGCGCGAGAAAGATGATGTTGTTAATGGCTT 76565
109 rIleuProAsnArgAsnIlePheArgTyrIleThrArgGlnSer 126
76564 TTCCTCGCGGGAATATTTTATGTTTAAAGCGGAAACGAGAGCTT 76515
126 eHisSerLeuSerProPheMetAspAspPheValProGlyValAsn 142
76514 TGAATTTGATCCGCTTTT...GATTCGATGCTGTAGTGGTGTAC 76468
143 HisSerProValIleAlaProArgLysValProArgSerProTyrLys 158
76467 CCTAGTCTGTTAAGTCGCCGAGAAAGATCTTAGGCTCTTAAAGGT 76418
158 ..... 158
76417 CAGTGTGTTTGTATCTTCGGGGTCTTGTGGAGTTGAAATTAG 76368
158 ..... 158
76367 GTTGTAGATGTTGTTCTTGATGATTTGCTATGTTGGTAGATT 76318
158 ..... 158
76317 TAGGTTTGGAGCTGTGATCTTGATGTTGTTGCACATTTATCTA 76268
158 ..... 158
76267 TTTGTCATTTATGATGAGACTCTTGTCTAGTGGCTGATGTT 76218
158 ..... 158
76217 TTGAAACGTCTAGTGTGTCATTTTCTACTCTTTATCCAGTT 76168
158 ..... 158
76167 TGAATTAGAAATGTTTCTTTTGTGCATTTTGCCTGTGATGAGCTG 76118
159 ..... 158
76117 AAAAGCTTCTTTTGTGTTTATGATGATGATGATGATGATGATG 76068
168 spPheTyrLeuAsnLeuValAspTyrPheSerHisAsnValLeuAlaVal 184
76067 ATTTTACTTGAATCTCGTGAATGTCGCGACAAATGTTCTTCAAGTG 76018
185 GlyLeuGlyAsnCysValTyrLeuTyrAsnAlaCysSerSerLys..Val 200
76017 GGATTAGGAACTCGCTGTATTTTGTGGAATGCTTGTAGCAAGGTGAG 75968
200 ..... 200
75967 CTCTTTTCTTCTGATGATGATGATGATGATGATGATGATGAT 75918
201 ..... 75918
75917 AGTGTCTTTGTATGAAATCTCTTCTGCTATTTGAGTAAGTACTAGT 75868
ThrLysL 203
|||||
75868
203 euCysAspLeuGlyValAspAspCysValCysSerValGlyTyrAlaGln 219
75867 TATGCAATCTTGGGGTGCATGAAACCTGTTGCTCAGTTGTTGGCACTA 75818
220 ArgGlyThrHisLeuAlaValGlyThrAsnAsnGlyLysVal..... 233
75817 CGTGAACACATTTGGCTATTTGCACTAGTACCGGACAGTACAGTTAG 75768
233 ..... 233
75767 AATATCAATGTTTGTGCTATCTGAGTTATATCATGCTTATATATA 75718
233 ..... 233
75717 GTACTCTTGAATCTTGAATCGGAGAGATCTTTTGGAAACTGATT 75668
233 ..... 233
75667 GAATTAGACTTTGAGATGTTGCTTTTGCATAATGTTATTCATGAA 75618
233 ..... 233
75617 GCATTGCTAGCTTACTTTCACCAATCATATGACTCATTAATGATC 75568
233 ..... 233
75567 GTTGGTTTGAAGATGATATTTGCTTAAACACCAACAGATCAAGTAA 75518
233 ..... 233
75517 GCAGCCATCTGATCTTCTTAAAGGTGATGATCTTATAGAGAA 75468
234 ..... 234
75467 CTTACGTTACTGATTTCTTAAATGATGATGATGATGATGATGATG 75418
242 sLysIleArgSerMetGluGlyHisArgLeuValGlyAlaLeuAla 259
75417 GAATATTGAACAAATGAGGATCGCTAAGGTTGGAGCTTGGCAT 75368
259 rPheSerSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGln 275
75367 GGAGCTCATGTTCTGCTCTGAGTAGAGACAAAGACATCTTACG 75318
276 ArgAspIleArgTyrGlnGluAspPheValSerLysLeuSerGlyHisLys 292
75317 AGAGCATGTTGACTCAAGAAATCATGTCAGTAACTTAAAGTCAACA 75268
292 sSerGluValCysGlyLeuLysTyrPheTyrAspAsnArgGluLeuAla 309
75267 ATCCGAATCTGTGACTCAAAATGCTTCCGCAATGCGAGCTAGCAT 75218
309 eArgGlyAsnAspAsnLys..... 315
75217 CAGGTGAACGACATTAAGTAATATATATATCATCTCTGTCTGGA 75168
316 ..... 316
75167 AGTTAACATGAATTTTCAACATATGATGATCTAGATCTTGCAGCT 75118
316 upheValTyrAsnGlnHisSerThrGlnProValLeuLysTyrCysGlu 333
75117 GTTGTATGAAACCAATTCACACACACGCTTGTGATCTGTAAC 75068
333 iThrAlaAlaValAlaLeuAlaTyrPheProHisLeuHisGlyLeu 349
75067 ACGCCGACAGAGTGAAGCCATTCGTGCTCACATGATTTTGGACTT 75018
350 LeuAlaSerGlyGlyLysThrAlaAspArgCysIleArgPheTyrAsn 366
75017 CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 74968
366 rThrThrAsnSerHisLeuSerCysMetAspThrGlySerGlnVal.... 381

```

```

|||||.....|||||.....|||||.....
74967 AACGACAAACACTCATTTAAATTGGTAGATCCAAACAGCAAGTAATC 74918
381 ..... 381
74917 CGGTATGCTACTATTAACATCTGAGATTTCACATGTGGTTTATAG 74868
382 .....Cys..AsnLeuValTrpSerLys 388
74867 CAGCAAAATCTCTTCCACACGACGAGGTGTGAATTTGGTAGCTAAG 74818
389 AsnValAsnGluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleI 405
|||||.....|||||.....|||||.....
74817 AACGTGAATGACCTTGTAGACACACGCGGTATTCACCAACCAATCAT 74768
405 eValTrpArgTyrProThrMetSerLys..... 414
|||||.....|||||.....|||||.....
74767 CCTCTGGAATATCCACACCATGTCTAAAGTAAGATTTCTTCTTACT 74718
414 ..... 414
74717 TCGTTTTTTTATGTGCTGATTTGGATTTTGATGAGAACCTGTT 74668
415 .....LeuAlaThrLeuThrGlyHisThrTyrArgValLeuTyrIle 428
74667 CAGTATATGTTGGCACTCTGCTGCTACCTGACCTGACCGCGTTGTACT 74618
428 uAlaIleSerProAspGly.Gln..... 435
|||||.....|||||.....
74617 TCGGCTGTCACCGGATGACAGAGTACTTTTCTTCTCTGCTGCTC 74568
435 ..... 435
74567 ATTCTTAATCATCATCTATATTTCCTTCTTAAACACTTG 74518
436 .....ThrIleValIleArgIleIle 442
74517 CATGACACACTTCTTCTGATTTGTCAGACAACTGTGACGAGAGCAG 74468
442 lValAspGluThrLeuArgPheThrAsnValIlePheProSerProLys..Ser 457
|||||.....|||||.....|||||.....
74467 GAGATGAACCTGTGAGATTCTGAGATGCTCCCTTCGCAAAATCTCAG 74418
457 ..... 457
74417 GTAAAGAACCCGAGCATATTCGCCATTTCTAATCCCAATTTCTTGA 74368
457 ..... 457
74367 AATGACAAATGATTAACACACACTTCTTTTCTTTATTTTATTTT 74318
458 .....GlnAsnThrGluSerGluIleGlyAlaLeuSerLeuGlyArgT 472
|||||.....|||||.....|||||.....
74317 CATATCTGACAGCAGGAGAGCAAAATCGGGCATTAATCTTTGTAGAA 74268
472 hTThrIleArg 475
|||||.....
74267 CAACAATCCGG 74257
seq_name: gb_pl:ATCHRIV33
seq_documentation_block: 190026 bp DNA linear, PLN 16-MAR-2000
LOCUS ATCHRIV33
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 33.
ACCESSION AL161533
VERSION AL161533.2 GI:7267889
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 2832 to 3389; 3666 to 4213)

```

```

AUTHORS Volckaert, G., Gijmenez, B., Voet, M., Robben, J., Mewes, H.W.,
JOURNAL Lemcke, K. and Mayer, K.F.X.
REFERENCE Unpublished
2 (bases 123056 to 190026)
AUTHORS Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W.,
JOURNAL Lemcke, K. and Mayer, K.F.X.
REFERENCE Unpublished
3 (bases 1 to 34336; 23831 to 131430)
AUTHORS Hilbert, H., Braun, M., Holzer, E., Brandt, A., Duesterhoeft, A.,
JOURNAL Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
REFERENCE Unpublished
4 (bases 1 to 190026)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, proj/thal/
Coordinator: Mike Berran, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK.
E-mail: michael.berran@bsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV32 at the 5' end and an
overlap with ATCHRIV34 at the 3' end.
FEATURES
source Location/Qualifiers
1..190026
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
5417..6920
/gene="AT4g11890"
5417..5486
/gene="AT4g11890"
/number=1
join(5417..5486,5584..5675,5763..5973,6065..6302,
6396..6534,6612..6920)
/gene="AT4g11890"
/number=1
/number=2
/number=3
/number=4
/number=5
/number=6
/number=7
/number=8
/number=9
/number=10
/number=11
/number=12
/number=13
/number=14
/number=15
/number=16
/number=17
/number=18
/number=19
/number=20
/number=21
/number=22
/number=23
/number=24
/number=25
/number=26
/number=27
/number=28
/number=29
/number=30
/number=31
/number=32
/number=33
/number=34
/number=35
/number=36
/number=37
/number=38
/number=39
/number=40
/number=41
/number=42
/number=43
/number=44
/number=45
/number=46
/number=47
/number=48
/number=49
/number=50
/number=51
/number=52
/number=53
/number=54
/number=55
/number=56
/number=57
/number=58
/number=59
/number=60
/number=61
/number=62
/number=63
/number=64
/number=65
/number=66
/number=67
/number=68
/number=69
/number=70
/number=71
/number=72
/number=73
/number=74
/number=75
/number=76
/number=77
/number=78
/number=79
/number=80
/number=81
/number=82
/number=83
/number=84
/number=85
/number=86
/number=87
/number=88
/number=89
/number=90
/number=91
/number=92
/number=93
/number=94
/number=95
/number=96
/number=97
/number=98
/number=99
/number=100
/number=101
/number=102
/number=103
/number=104
/number=105
/number=106
/number=107
/number=108
/number=109
/number=110
/number=111
/number=112
/number=113
/number=114
/number=115
/number=116
/number=117
/number=118
/number=119
/number=120
/number=121
/number=122
/number=123
/number=124
/number=125
/number=126
/number=127
/number=128
/number=129
/number=130
/number=131
/number=132
/number=133
/number=134
/number=135
/number=136
/number=137
/number=138
/number=139
/number=140
/number=141
/number=142
/number=143
/number=144
/number=145
/number=146
/number=147
/number=148
/number=149
/number=150
/number=151
/number=152
/number=153
/number=154
/number=155
/number=156
/number=157
/number=158
/number=159
/number=160
/number=161
/number=162
/number=163
/number=164
/number=165
/number=166
/number=167
/number=168
/number=169
/number=170
/number=171
/number=172
/number=173
/number=174
/number=175
/number=176
/number=177
/number=178
/number=179
/number=180
/number=181
/number=182
/number=183
/number=184
/number=185
/number=186
/number=187
/number=188
/number=189
/number=190
/number=191
/number=192
/number=193
/number=194
/number=195
/number=196
/number=197
/number=198
/number=199
/number=200
/number=201
/number=202
/number=203
/number=204
/number=205
/number=206
/number=207
/number=208
/number=209
/number=210
/number=211
/number=212
/number=213
/number=214
/number=215
/number=216
/number=217
/number=218
/number=219
/number=220
/number=221
/number=222
/number=223
/number=224
/number=225
/number=226
/number=227
/number=228
/number=229
/number=230
/number=231
/number=232
/number=233
/number=234
/number=235
/number=236
/number=237
/number=238
/number=239
/number=240
/number=241
/number=242
/number=243
/number=244
/number=245
/number=246
/number=247
/number=248
/number=249
/number=250
/number=251
/number=252
/number=253
/number=254
/number=255
/number=256
/number=257
/number=258
/number=259
/number=260
/number=261
/number=262
/number=263
/number=264
/number=265
/number=266
/number=267
/number=268
/number=269
/number=270
/number=271
/number=272
/number=273
/number=274
/number=275
/number=276
/number=277
/number=278
/number=279
/number=280
/number=281
/number=282
/number=283
/number=284
/number=285
/number=286
/number=287
/number=288
/number=289
/number=290
/number=291
/number=292
/number=293
/number=294
/number=295
/number=296
/number=297
/number=298
/number=299
/number=300
/number=301
/number=302
/number=303
/number=304
/number=305
/number=306
/number=307
/number=308
/number=309
/number=310
/number=311
/number=312
/number=313
/number=314
/number=315
/number=316
/number=317
/number=318
/number=319
/number=320
/number=321
/number=322
/number=323
/number=324
/number=325
/number=326
/number=327
/number=328
/number=329
/number=330
/number=331
/number=332
/number=333
/number=334
/number=335
/number=336
/number=337
/number=338
/number=339
/number=340
/number=341
/number=342
/number=343
/number=344
/number=345
/number=346
/number=347
/number=348
/number=349
/number=350
/number=351
/number=352
/number=353
/number=354
/number=355
/number=356
/number=357
/number=358
/number=359
/number=360
/number=361
/number=362
/number=363
/number=364
/number=365
/number=366
/number=367
/number=368
/number=369
/number=370
/number=371
/number=372
/number=373
/number=374
/number=375
/number=376
/number=377
/number=378
/number=379
/number=380
/number=381
/number=382
/number=383
/number=384
/number=385
/number=386
/number=387
/number=388
/number=389
/number=390
/number=391
/number=392
/number=393
/number=394
/number=395
/number=396
/number=397
/number=398
/number=399
/number=400
/number=401
/number=402
/number=403
/number=404
/number=405
/number=406
/number=407
/number=408
/number=409
/number=410
/number=411
/number=412
/number=413
/number=414
/number=415
/number=416
/number=417
/number=418
/number=419
/number=420
/number=421
/number=422
/number=423
/number=424
/number=425
/number=426
/number=427
/number=428
/number=429
/number=430
/number=431
/number=432
/number=433
/number=434
/number=435
/number=436
/number=437
/number=438
/number=439
/number=440
/number=441
/number=442
/number=443
/number=444
/number=445
/number=446
/number=447
/number=448
/number=449
/number=450
/number=451
/number=452
/number=453
/number=454
/number=455
/number=456
/number=457
/number=458
/number=459
/number=460
/number=461
/number=462
/number=463
/number=464
/number=465
/number=466
/number=467
/number=468
/number=469
/number=470
/number=471
/number=472
/number=473
/number=474
/number=475
/number=476
/number=477
/number=478
/number=479
/number=480
/number=481
/number=482
/number=483
/number=484
/number=485
/number=486
/number=487
/number=488
/number=489
/number=490
/number=491
/number=492
/number=493
/number=494
/number=495
/number=496
/number=497
/number=498
/number=499
/number=500
/number=501
/number=502
/number=503
/number=504
/number=505
/number=506
/number=507
/number=508
/number=509
/number=510
/number=511
/number=512
/number=513
/number=514
/number=515
/number=516
/number=517
/number=518
/number=519
/number=520
/number=521
/number=522
/number=523
/number=524
/number=525
/number=526
/number=527
/number=528
/number=529
/number=530
/number=531
/number=532
/number=533
/number=534
/number=535
/number=536
/number=537
/number=538
/number=539
/number=540
/number=541
/number=542
/number=543
/number=544
/number=545
/number=546
/number=547
/number=548
/number=549
/number=550
/number=551
/number=552
/number=553
/number=554
/number=555
/number=556
/number=557
/number=558
/number=559
/number=560
/number=561
/number=562
/number=563
/number=564
/number=565
/number=566
/number=567
/number=568
/number=569
/number=570
/number=571
/number=572
/number=573
/number=574
/number=575
/number=576
/number=577
/number=578
/number=579
/number=580
/number=581
/number=582
/number=583
/number=584
/number=585
/number=586
/number=587
/number=588
/number=589
/number=590
/number=591
/number=592
/number=593
/number=594
/number=595
/number=596
/number=597
/number=598
/number=599
/number=600
/number=601
/number=602
/number=603
/number=604
/number=605
/number=606
/number=607
/number=608
/number=609
/number=610
/number=611
/number=612
/number=613
/number=614
/number=615
/number=616
/number=617
/number=618
/number=619
/number=620
/number=621
/number=622
/number=623
/number=624
/number=625
/number=626
/number=627
/number=628
/number=629
/number=630
/number=631
/number=632
/number=633
/number=634
/number=635
/number=636
/number=637
/number=638
/number=639
/number=640
/number=641
/number=642
/number=643
/number=644
/number=645
/number=646
/number=647
/number=648
/number=649
/number=650
/number=651
/number=652
/number=653
/number=654
/number=655
/number=656
/number=657
/number=658
/number=659
/number=660
/number=661
/number=662
/number=663
/number=664
/number=665
/number=666
/number=667
/number=668
/number=669
/number=670
/number=671
/number=672
/number=673
/number=674
/number=675
/number=676
/number=677
/number=678
/number=679
/number=680
/number=681
/number=682
/number=683
/number=684
/number=685
/number=686
/number=687
/number=688
/number=689
/number=690
/number=691
/number=692
/number=693
/number=694
/number=695
/number=696
/number=697
/number=698
/number=699
/number=700
/number=701
/number=702
/number=703
/number=704
/number=705
/number=706
/number=707
/number=708
/number=709
/number=710
/number=711
/number=712
/number=713
/number=714
/number=715
/number=716
/number=717
/number=718
/number=719
/number=720
/number=721
/number=722
/number=723
/number=724
/number=725
/number=726
/number=727
/number=728
/number=729
/number=730
/number=731
/number=732
/number=733
/number=734
/number=735
/number=736
/number=737
/number=738
/number=739
/number=740
/number=741
/number=742
/number=743
/number=744
/number=745
/number=746
/number=747
/number=748
/number=749
/number=750
/number=751
/number=752
/number=753
/number=754
/number=755
/number=756
/number=757
/number=758
/number=759
/number=760
/number=761
/number=762
/number=763
/number=764
/number=765
/number=766
/number=767
/number=768
/number=769
/number=770
/number=771
/number=772
/number=773
/number=774
/number=775
/number=776
/number=777
/number=778
/number=779
/number=780
/number=781
/number=782
/number=783
/number=784
/number=785
/number=786
/number=787
/number=788
/number=789
/number=790
/number=791
/number=792
/number=793
/number=794
/number=795
/number=796
/number=797
/number=798
/number=799
/number=800
/number=801
/number=802
/number=803
/number=804
/number=805
/number=806
/number=807
/number=808
/number=809
/number=810
/number=811
/number=812
/number=813
/number=814
/number=815
/number=816
/number=817
/number=818
/number=819
/number=820
/number=821
/number=822
/number=823
/number=824
/number=825
/number=826
/number=827
/number=828
/number=829
/number=830
/number=831
/number=832
/number=833
/number=834
/number=835
/number=836
/number=837
/number=838
/number=839
/number=840
/number=841
/number=842
/number=843
/number=844
/number=845
/number=846
/number=847
/number=848
/number=849
/number=850
/number=851
/number=852
/number=853
/number=854
/number=855
/number=856
/number=857
/number=858
/number=859
/number=860
/number=861
/number=862
/number=863
/number=864
/number=865
/number=866
/number=867
/number=868
/number=869
/number=870
/number=871
/number=872
/number=873
/number=874
/number=875
/number=876
/number=877
/number=878
/number=879
/number=880
/number=881
/number=882
/number=883
/number=884
/number=885
/number=886
/number=887
/number=888
/number=889
/number=890
/number=891
/number=892
/number=893
/number=894
/number=895
/number=896
/number=897
/number=898
/number=899
/number=900
/number=901
/number=902
/number=903
/number=904
/number=905
/number=906
/number=907
/number=908
/number=909
/number=910
/number=911
/number=912
/number=913
/number=914
/number=915
/number=916
/number=917
/number=918
/number=919
/number=920
/number=921
/number=922
/number=923
/number=924
/number=925
/number=926
/number=927
/number=928
/number=929
/number=930
/number=931
/number=932
/number=933
/number=934
/number=935
/number=936
/number=937
/number=938
/number=939
/number=940
/number=941
/number=942
/number=943
/number=944
/number=945
/number=946
/number=947
/number=948
/number=949
/number=950
/number=951
/number=952
/number=953
/number=954
/number=955
/number=956
/number=957
/number=958
/number=959
/number=960
/number=961
/number=962
/number=963
/number=964
/number=965
/number=966
/number=967
/number=968
/number=969
/number=970
/number=971
/number=972
/number=973
/number=974
/number=975
/number=976
/number=977
/number=978
/number=979
/number=980
/number=981
/number=982
/number=983
/number=984
/number=985
/number=986
/number=987
/number=988
/number=989
/number=990
/number=991
/number=992
/number=993
/number=994
/number=995
/number=996
/number=997
/number=998
/number=999
/number=1000
/number=1001
/number=1002
/number=1003
/number=1004
/number=1005
/number=1006
/number=1007
/number=1008
/number=1009
/number=1010
/number=1011
/number=1012
/number=1013
/number=1014
/number=1015
/number=1016
/number=1017
/number=1018
/number=1019
/number=1020
/number=1021
/number=1022
/number=1023
/number=1024
/number=1025
/number=1026
/number=1027
/number=1028
/number=1029
/number=1030
/number=1031
/number=1032
/number=1033
/number=1034
/number=1035
/number=1036
/number=1037
/number=1038
/number=1039
/number=1040
/number=1041
/number=1042
/number=1043
/number=1044
/number=1045
/number=1046
/number=1047
/number=1048
/number=1049
/number=1050
/number=1051
/number=1052
/number=1053
/number=1054
/number=1055
/number=1056
/number=1057
/number=1058
/number=1059
/number=1060
/number=1061
/number=1062
/number=1063
/number=1064
/number=1065
/number=1066
/number=1067
/number=1068
/number=1069
/number=1070
/number=1071
/number=1072
/number=1073
/number=1074
/number=1075
/number=1076
/number=1077
/number=1078
/number=1079
/number=1080
/number=1081
/number=1082
/number=1083
/number=1084
/number=1085
/number=1086
/number=1087
/number=1088
/number=1089
/number=1090
/number=1091
/number=1092
/number=1093
/number=1094
/number=1095
/number=1096
/number=1097
/number=1098
/number=1099
/number=1100
/number=1101
/number=1102
/number=1103
/number=1104
/number=1105
/number=1106
/number=1107
/number=1108
/number=1109
/number=1110
/number=1111
/number=1112
/number=1113
/number=1114
/number=1115
/number=1116
/number=1117
/number=1118
/number=1119
/number=1120
/number=1121
/number=1122
/number=1123
/number=1124
/number=1125
/number=1126
/number=1127
/number=1128
/number=1129
/number=1130
/number=1131
/number=1132
/number=1133
/number=1134
/number=1135
/number=1136
/number=1137
/number=1138
/number=1139
/number=1140
/number=1141
/number=1142
/number=1143
/number=1144
/number=1145
/number=1146
/number=1147
/number=1148
/number=1149
/number=1150
/number=1151
/number=1152
/number=1153
/number=1154
/number=1155
/number=1156
/number=1157
/number=1158
/number=1159
/number=1160
/number=1161
/number=1162
/number=1163
/number=1164
/number=1165
/number=1166
/number=1167
/number=1168
/number=1169
/number=1170
/number=1171
/number=1172
/number=1173
/number=1174
/number=1175
/number=1176
/number=1177
/number=1178
/number=1179
/number=1180
/number=1181
/number=1182
/number=1183
/number=1184
/number=1185
/number=1186
/number=1187
/number=1188
/number=1189
/number=1190
/number=1191
/number=1192
/number=1193
/number=1194
/number=1195
/number=1196
/number=1197
/number=1198
/number=1199
/number=1200
/number=1201
/number=1202
/number=1203
/number=1204
/number=1205
/number=1206
/number=1207
/number=1208
/number=1209
/number=1210
/number=1211
/number=1212
/number=1213
/number=1214
/number=1215
/number=1216
/number=1217
/number=1218
/number=1219
/number=1220
/number=1221
/number=1222
/number=1223
/number=1224
/number=1225
/number=1226
/number=1227
/number=1228
/number=1229
/number=1230
/number=1231
/number=1232
/number=1233
/number=1234
/number=1235
/number=1236
/number=1237
/number=1238
/number=1239
/number=1240
/number=1241
/number=1242
/number=1243
/number=1244
/number=1245
/number=1246
/number=1247
/number=1248
/number=1249
/number=1250
/number=1251
/number=1252
/number=1253
/number=1254
/number=1255
/number=1256
/number=1257
/number=1258
/number=1259
/number=1260
/number=1261
/number=1262
/number=1263
/number=1264
/number=1265
/number=1266
/number=1267
/number=1268
/number=1269
/number=1270
/number=1271
/number=1272
/number=1273
/number=1274
/number=1275
/number=1276
/number=1277
/number=1278
/number=1279
/number=1280
/number=1281
/number=1282
/number=1283
/number=1284
/number=1285
/number=1286
/number=1287
/number=1288
/number=1289
/number=1290
/number=1291
/number=1292
/number=1293
/number=1294
/number=1295
/number=1296
/number=1297
/number=1298
/number=1299
/number=1300
/number=1301
/number=1302
/number=1303
/number=1304
/number=1305
/number=1306
/number=1307
/number=1308
/number=1309
/number=1310
/number=1311
/number=1312
/number=1313
/number=1314
/number=1315
/number=1316
/number=1317
/number=1318
/number=1319
/number=1320
/number=1321
/number=1322
/number=1323
/number=1324
/number=1325
/number=1326
/number=1327
/number=1328
/number=1329
/number=1330
/number=1331
/number=1332
/number=1333
/number=1334
/number=1335
/number=1336
/number=1337
/number=1338
/number=1339
/number=1340
/number=1341
/number=1342
/number=1343
/number=1344
/number=1345
/number=1346
/number=1347
/number=1348
/number=1349
/number=1350
/number=1351
/number=1352
/number=1353
/number=1354
/number=1355
/number=1356
/number=1357
/number=1358
/number=1359
/number=1360
/number=1361
/number=1362
/number=1363
/number=1364
/number=1365
/number=1366
/number=1367
/number=1368
/number=1369
/number=1370
/number=1371
/number=1372
/number=1373
/number=1374
/number=1375
/number=1376
/number=1377
/number=1378
/number=1379
/number=1380
/number=1381
/number=1382
/number=1383
/number=1384
/number=1385
/number=1386
/number=1387
/number=1388
/number=1389
/number=1390
/number=1391
/number=1392
/number=1393
/number=1394
/number=1395
/number=1396
/number=1397
/number=1398
/number=1399
/number=1400
/number=1401
/number=1402
/number=1403
/number=1404
/number=1405
/number=1406
/number=1407
/number=1408
/number=1409
/number=1410
/number=1411
/number=1412
/number=1413
/number=1414
/number=1415
/number=1416
/number=1417
/number=1418
/number=1419
/number=1420
/number=1421
/number=1422
/number=1423
/number=1424
/number=1425
/number=1426
/number=1427
/number=1428
/number=1429
/number=1430
/number=1431
/number=1432
/number=1433
/number=1434
/number=1435
/number=1436
/number=1437
/number=1438
/number=1439
/number=1440
/number=1441
/number=1442
/number=1443
/
```

```

/gene="AT4g11890"
/number=4
6396..6534
/gene="AT4g11890"
/number=5
6535..6611
/gene="AT4g11890"
/number=5
6612..6920
/gene="AT4g11890"
/number=6
7389..10690
/gene="AT4g11900"
complement(7389..7679)
/gene="AT4g11900"
/number=1
complement(join(7389..7679,7778..7928,8208..8445,
8355..8765,8889..9043,9121..9255,9322..10690))
/gene="AT4g11900"
complement(join(7389..7679,7778..7928,8208..8445,
8555..8765,8889..9043,9121..9255,9322..10690))
/gene="AT4g11900"
/note="strong similarity to KI domain interacting kinase 1
- Zea mays, PID:g2735017
Contains Prokaryotic membrane lipoprotein lipid attachment
site AA472-482; protein kinases signatures and profile
AA658-670
contains EST gb:AT992820.1, T20644, AA586217"
/codon_start=1
/product="KI domain interacting kinase 1-like protein"
/protein_id="CAB78233.1"
/db_xref="GI:7267891"
/translation="MQICKKNVFLYYGVLPSPQVSSSDTISTNOPLSGFETIVS
SGIFELGFTPTDPTDHRNYIGMYRHVGPOTIVVANNRSPDAGGAFYLLKIL
DGNLILHNDISATRSKSTEGTSRSPQKISEGILFHEHWGSSKQVOAVLF
ISTANNRKTESKSCIVPLVLAIVAAACEVGLGCCISSRIRKKRORREKRE
LLEGLLIDAGEMCYLNDIMVATNSPFRKKLGESEGPVYKQKLPNGEVAIKR
LSKRSQGLTERKNEVYLITKLOHKNVYLLGCVBGDKLLIYETMSKSIDGLIFP
SLKSRLEDMETRRKIVNGTTRGLQYHESRLRIIHRDLKASNLILDDMNKRLIDFG
TARIFCKQIDISTORIVGTFGMSPEVYALGVISSEKSDIYSEFVLLLEIIIGKQATR
FVNDQKHSLIAYEWSMCTEGVSIIDPCCSGLSEAMCIIHALICVDHFKDR
PMISQIVYMLNDNTLPIKOPFESNVNLNGDQDLVVFISINATQTELEAR"
complement(7680..7777)
/gene="AT4g11900"
/number=1
complement(7778..7928)
/gene="AT4g11900"
/number=2
complement(7929..8207)
/gene="AT4g11900"
/number=2
complement(8208..8445)
/gene="AT4g11900"
/number=3
complement(8446..8554)
/gene="AT4g11900"
/number=3
complement(8555..8765)
/gene="AT4g11900"
/number=4
complement(8766..8888)
/gene="AT4g11900"
/number=4
complement(8889..9043)
/gene="AT4g11900"
/number=5
complement(9044..9120)
/gene="AT4g11900"

```

```

/number=5
complement(9121..9255)
/gene="AT4g11900"
/number=6
complement(9256..9321)
/gene="AT4g11900"
/number=6
complement(9322..10690)
/gene="AT4g11900"
/number=7
13583..16588
/gene="AT4g11910"
13583..13711
/gene="AT4g11910"
/number=1
join(13583..13711,13978..14151,14239..14406,14643..14879,
15546..15719,15829..15996,16338..16588)
/gene="AT4g11910"
/note="similarity to hypothetical protein F7H19.100
-Arabidopsis thaliana, PID:e1310060
contains zinc finger, C2H2 type, domain AA420-440
contains EST gb:Z34017, Z34018"
/codon_start=1
/product="putative protein"
/protein_id="CAB78234.1"
/db_xref="GI:7267892"
/translation="MCSLATMLLPSSKMKPVPEPKLSTSSLCVYTRSKKNSIVPV
ARLPGALFEASKLKLVLGVDKHHKAPKLPRTYTLHSDITAKLTLAISQISINSOL
OGMANKLPREDVEGMKKVKMSLHVCHISGHEFLNLAKRLRYIFCKELPVLE
APANGDEYLLNNHPELOESPVYVYFHNKVECGMPLMEASOHODGTHRS
ETLPELPCPDCKCCFPYVNLFGQALFEASKLVKFLGVDEKHPNLPRTYTHS
DITAKLTLAISQISINSOLOGMANKRLRDEVAEMKVKSMSLHVCHISGHEFLD
LIAELRYEIFCKELPMVLKAVHGDENMLNLYPELHAFVYVYFHNIPKNNVECG
RLCEATSIDGCKPTPCETLPEPCFDKSCCFPVSTIPWSHSGCSHGEDNNVALA
GNLITKMYRSRKK"
13712..13977
/gene="AT4g11910"
/number=1
13978..14151
/gene="AT4g11910"
/number=2
14152..14238
/gene="AT4g11910"

```

alignment_scores:

Quality:	1519.50	Length:	889
Ratio:	3.559	Gaps:	19
Percent Similarity:	48.031	Percent Identity:	40.607

alignment_block:

US-09-701-572-2 x ATCHRIV33/rev ..

Align seg 1/1 to reverse of: ATCHRIV33 from: 1 to: 190026

```

1 MetAspGlyThrGlyAsnArgAnProProthSerThrValArgAs 17
||||| : : : : : ||| : :
20405 ATGGAAGAGATGATCAACACACGAGAAAGAACTCAGCT 20356
17 pAsnSerProPro.....ProGluProSerProGluSerLeuA 30
||| ||||| ||| ||| |||||
20355 GAATCTTCACCGCTGATGATGTCGACAGGTCACACGATC... 20310
30 rGHisValSerArgMetIleAsnSerAsnHisTyrThrSerProSerArg 46
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
20309 .CGAATCAACCGTTAATGATTCGAATCAATATCATCTCTCCAGAA 20262
47 ThrIleTyrSerAspArgPheIleProSerArgSerAlaSerIlePheAl 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20261 CCTATTACTCAGATAGATTATTCAGATGATGCTGTTCCAAATTCGC 20212
63 AleuPheAspIle.....AsnThrProThr.....GluGlyArgAspA 76
||||| ||||| : : : : : : : : : : : : : : : : : : : : :
20211 GCTTTTCGATTTCGATCTTCTTCGCCGAATAGAAAGATGGGAAGAAG 20162

```



```

18379 CTGCATATCTTCTTCCACAGCAGCTGTCAATTGGTATGCTTAAG 18330
389 AasnValAsnGluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleI 405
18329 AACGTAAAGACCTGTGTAGACACACGGGTATTCACCAACCAATATCAT 18280
405 eValTPaGlyTyrProThrMetSerLys..... 414
18279 CGCTCGAATAATATCCACCATGTCTAAGTAATCTTCTCTTACT 18230
414 ..... 414
18229 TCGTTTTTATGTGCTGTAATTTGGATTTTGATGAGACCTGTT 18180
415 ..... LeuAlaThrLeuThrGlyHisThrTyrArgValLeuTyrLe 428
18179 CAGTATTAAGTGGCAACTCTGCTACTGCTACCTGACCGGCTTCTACT 18130
428 uAlaIleSerProAspGly.Gln..... 435
18129 TCGGGGTACCGGAGAGAGAGAGGTTACTTTTCTTCTCTCTGCTC 18080
435 ..... 435
18079 ATTCTTCTAATCATCATCTTATATTTTGGCTACTTCTTATACACTTG 18030
436 ..... ThrIleValThrGlyAlaG 442
18029 CATCAACACACTTCTTCTTATTTGTCGACAAATTTGACAGAGAGAG 17980
442 lYAspIuThrLeuArgPheThrPasnValPheProSerProLys..Ser. 457
17979 GAGATGAACCTTGAGATTCTGGAATGCTTCCCTTCCGGAATCTCAG 17930
457 ..... 457
17929 GTAAAGAACCCGAGACATATTCGCCATTTCTAATCTCAATTTCTTGA 17880
457 ..... 457
17879 AATGACAACTGATTAACACACAACTCTTCTTCTTATTTT 17830
458 ..... GlnAsnThrGluSerGluIleGlyAlaLeuSerLeuGlyArgT 472
17829 CTATCTCTGAGCAGGAGAGAGGAATCGGGCATTAATCTTGTGTAGA 17780
472 hTThrIleArg 475
17779 CAACAATCCGG 17769

seq_name: gb_B013413
seq_documentation_block:
LOCUS BC013413 1866 bp mRNA linear PRI 04-SEP-2001
DEFINITION Homo sapiens, clone MGC:4589 IMAGE:3160334, mRNA, complete cds.
ACCESSION BC013413
VERSION BC013413.1 GI:15426580
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1866)
Strausberg, R.
Direct Submission
Submitted (31-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NHI-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-femail.nih.gov

```

```

FEATURES
    source
        1..1866
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="MGC:4589 IMAGE:3160334"
            /tissue_type="Brain, neuroblastoma"
            /clone_lib="NIH_MGC_19"
            /lab_host="DH10B-R"
            /note="Vector: pOTB7"
            /codon_start=1
            /product="Unknown (protein for MGC:4589)"
            /protein_id="AAH13413.1"
            /db_xref="gi:15426581"
            /translation="MDQYERLRLOIVQIENMPRTREKRLTPASSPVSPKSH
            GREFPSRAGAWSVNFRINENKSPQNKRAADATSDNKDLAVALKKNLLDA
            GLEKVDQPTEDRRLQSTPEKGLFTYSLSTRSSPDGNDVSPYSNKSORTL
            LSPRPKRIKIPKLEKVLDAPELDQDEYLVNVMSSLVNLSVGLTCVYLSACTSQ
            VTRLDLVEGDSVTSVGSSEKGNVAVGTGKGVQIWDAAAGKLSNLEGTARVGA
            LAWNEOLSSGRSDMLIQRIRTPRPGVSEKRGHQENCGKLKWSRHDOLASGND
            NKLIVNHSLSIPVOQYTEHLAAKATKMSRHQGLASGGGTRDRIRRNNTITGP
            LQCIDTGSQVNCILANSHANLVSHTGISQNIIVMTPSLTVAKLGHSTRLYLIA
            MSPDEALVTGAGDETLRFNMVFSKTRSKSVSLNLETRIR"

BASE COUNT      379 a      601 c      595 g      291 t
ORIGIN

alignment_scores:
    Quality: 1294.00      Length: 495
    Ratio: 3.469      Gaps: 14
    Percent Similarity: 75.354      Percent Identity: 54.343

alignment_block:
    US-09-701-572-2 x BC013413 ..
Align seg 1/1 to: BC013413 from: 1 to: 1866

6 AasnValAsnPropProThrSerThrValArgAspAsnSerPropPr 22
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
247 AATGACAAACGATGCCACCGCTCACAGAGATCGCGGACCTGACGCC 296
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
22 GlnProSerProGluSerLeuArgHisValSerArgMetIleAsnSer 39
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
297 TGCACGCTCCCAAGTGC..... 315
39 snHisTyrThrSerProSerArgThrIleTyrSerAspArgPheIlePro 55
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
316 .....TCGCCACGACAG.....CACGGAACCGCTTCATCACTCC 348
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
349 TCCAGAGCCGGAGACGACCTGAGCTGAACTTCCACAGATTACGAGAA 398
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
69 .....ThrProThrGluIleArg.....AspAspSerSerSer..... 79
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
399 TGAGAACTCTCCACAGTACAGAACGGAAGCAGAGACGCGACCTCAGACA 448
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
80 .....AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```


Quality:	1294.00	Length:	495
Ratio:	3.469	Gaps:	14
Percent Similarity:	75.354	Percent Identity:	54.343

```
alignment_block:
US-09-701-572-2 x AF083810 ..
```

Align seg 1/1 to: AF083810 from: 1 to: 2991

6 AsnArgHisProProProHisSerThrValAlaGlnAspAsnSerProPro 22
188 AATGAGAAACGATGGCAGCGGCACAGAGANTGGCGGACCCGAGGCC 237
22 OGIuProSerProGIuSerLeuArgHisValSerArgMetLeuSera 39
238 TGGCAGCTCCCCAGTGTCC..... 256
39 snHisTyrrHisSerProSerArgThrIleTyrrSerAspArgPheIlePro 55
257TGGCCAGAGAA.....CAGGGAGACGGCTTATGCC 289
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleasn..... 68
290 TCCAGACCCGAGCACAATGAGCGGTAACCTCCACAGAGATTACGAGAA 339
69ThrProThrgIuGIuArg.....AspAspSerSer.... 79
340 TGAGACAGTCCAGTCAACACACGAGAAAGCCAGAGCCACCTCAAGACA 389
80AlaTyrrThrIleLeuArgThrAlaLeuPhe 90
390 ACGGCAAAAGCGGCTGAGCTTACTGCTCCCTGTCTCAAGATGAGCTCTG 439
91 GLy.....ProAspValAlaGlyProValThrProGIuTyrrHis 104
440 GGTGCCGCGATCGAGAAAGTGGAGAGCCCGAGACTGAGACCGCAGCT 489
104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyrrLysThrG 121
490 GCAGCCCTCC...ACGCCTGAGAAAGAGGCTGTGTACGATCTCCCTTA 536
121 LuPThrArgInserMet.....HisSer 128
537 GCACACAGCGCTCCAGCCCGATGACGCGAACGATGTCTCCCTACTCC 586
129 LeuSerProPheMetLysAspAspArgPheValProGIuValAsnHisSerPr 145
587 CTGTCTCCGCTCAGCAAAAGAGC....CAGAAAGCTCGCCGCTGCC 630
145 oValLysAlaProArgLysValProArgSerProTyrrLysValLeuAspA 162
631 CCGGAAACCCACCCCGAAGANTCTCCAAAGTCCCTTAAGGTGCTGAGC 680
162 IaProAlaLeuGlnAspAspArgPheTyrrLeuAsnLeuValAspTPPserSer 178
681 CGGCCGAGCTGCAGAGCGACTTCTACCTCAATCTGTGGAGCTGTGCTGC 730
179 HisAsnValLeuAlaValAlaGIuLeuGIuAsnGlyAsnGlyValTyrrLeuTPsnaI 195
731 CTCAATGTGCTCAGCGTGGGCTAGCAGCACTGCTGATACCTGTGAGAGTCG 780
195 acySerSerLysValThrLysLeuGlyAspLeuGIuValAsp...AspC 211
781 CTGTACCAAGCCAGGTGACACGGCTCTGTGACCTCTCAGTGAAGGAGACT 830
211 ysaValGlySerValAlaGIyrrPAlaGlnArgGlyThrHisLeuAlaValGIy 227
831 CAATGACCTCCGCTGGCTGTGTGAGCGGGAACCTGAGCGGTGGCG 880
228 ThrAsnAsnGIyLysValGlnIlePrrAspAlaAlaArgCysLysTrI 244
881 ACACACAAAGGCTCTGTGACAGTCTGGGAGCAGACCCACGAGAAAGACT 930

244 eargsermetglnglyhisargleuthavalyalaleualatrpssr 261
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
931 gtcacattgttgagaggccacacgaccgcctggcgccgtcgtgatng 980
261 erSerLeuenserSerglyylarapslyasnlEtryGlnarasp 277
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
981 ctgaacacactctgccggagaccggcacccgccttgcgtcagagggac 1030
278 llaetgr.....GlngluapphevalserLyseuSerGlyHI 291
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1031 atccgacaccccgcacactgcagtccgmk.....CGCGGCTCCAGGGCCA 1074
291 slySerGlnValCysGlyLeuLysTrpSerTryraspaanargIuleua 308
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1075 ccggcagagaggtgtggcgacctcaagttggtccacagaccACAGCTCTGC 1124
308 laserGlylYasnAspasnlyleuPhevalTrpsnlnHisserThr 324
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1125 cctcgggggcgacacacacacactctgctgtcgaaatCACGCAAGCTGTG 1174
325 glnproVAlleuLysTryCysGlnHstHThAlaAlaVallysalalleal 341
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1175 agccccgtgcagcagtrAcAGGAGCAcCTGGCGGCCGTGAAGCCATCGC 1224
341 atrpserTrohisleuHisGlyleuLeualaserGlylglytrHAlaa 358
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1225 ctggcgtcccatCATCAGCAGGGTGTGGCTGTGGGGGGCGGCAAGCTG 1274
358 spArCyAllearPhetrPasnthrThrTrasnerHisLeuSerCys 374
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1275 accgctgATCCGCTTGAGACACCCTGACAAGAAACACTGGAGTGT 1324
375 MetaspthrGlySerGlnValCysAsnLeuValTrpSerLyASAnValas 391
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1335 ATCGACAGCGGCTCCCAAGTGTCAATCTGGCCTGTCCACACAGCCCA 1374
391 ngluenuvalserthrHisGlyTrsSerGlnasnGlnleileValTrpa 408
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1375 CGAGCTGTTGAGCACGACGACGGCTACTACAGAACCAGATCTTGTCTGGA 1424
408 rgyTrProthMetSerLyseuAlaThrleuThrGlyHisThrTyraG 424
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1425 AGTACCCCTCCCTGACCCAGGTGGCCAACTGCACGGGACCTCTACCGC 1474
425 ValleunryleuAlaHiserProaspyGlnThrHleAlaThyGlyAl 441
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1475 GTGCTGTACCTGGCAATGTCCCTGTATTGGGAGGCCATCTGACTGGTGGC 1524
441 agLYASPdlunThrleuarqPhetrrpasnalpheProserProlySerg 458
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1525 TGGAGAGAGACCTGAGGTTCGTGGAACGTCTT.....AGCAAACCC 1568
458 lnaenThGluserGluileGlyAlaLeuSerLeu 469
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1569 GTTCGACAAAAGAGTGTGTGTGTCTCAACCTC 1603

seq_name: gb_pr:AB033068

seq_documentation_block:

LOCUS AB033068 5008 bp mRNA linear PRI 11-NOV-1999

DEFINITION Homo sapiens mRNA for KIAA1242 protein, partial cds.

ACCESSION AB033068

VERSION AB033068.1 GI:6130804

KEYWORDS .

SOURCE Homo sapiens brain cDNA to mRNA, clone_1lb:pbluescriptII SK plus

clonename:h10406.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)

AUTHORS Nagase,T., Ishikawa,K., Kikuno,R., Hirosewa,M., Nomura,N. and Ohata,O.

TITLE Prediction of the coding sequences of unidentified human genes. XV.

The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
DNA Res. 6 (5), 337-345 (1999)

JOURNAL
MEDLINE
20039619
2 (bases 1 to 5008)
AUTHORS
Ohara, O., Nagase, T. and Kikuno, R.
TITLE
Direct Submission
JOURNAL

Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology, 1532-3 Yama, Kisarazu, Chiba
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,
URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)

FEATURES

Location/Qualifiers

1..5008
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="fh10406"
/issue_type="brain"
/clone_lib="pb1uescriptII SK plus"
1..1516
/gene="K1A1242"
1..1516
/gene="K1A1242"
/note="Start codon is not identified."
/product="K1A1242 protein"
/protein_id="BA86556.1"
/db_xref="GI:6330805"
/translation="LTLPRAPCLAMDQDERLLRQIVIONENTMPRTVETRRRLTP
ASSPSSPKHGDRFIPSRAGNMSVNFHRIENENESPKQNRKADATDNGKDLAY
SLIKELIAGIEKVEDQDPTDRRLQSTPKGLFTYSLTKSSPDGNDVSPYS
LSPVNSKSKLRLSPRKPRKISKIDPKYDLAPELLDDPFLNLVDMSSINLVSLIGT
CYLVNACTSYVTRICDLSVEGDSVTSVGSERGNLVANGTKGVQITDAAGKLS
MEGHARVAGALANNAEQLSSGSRDMILLQRIKTPPLQSERLQGHROEVGLKWT
DHOLLAGGNDKLLVNNHSLSPVOQTEHLAAVLAALMSRHQGLLASGGTQVADR
IREMVLTCGLOPLOCIDTGSQVCLAMSKHANEIVSYHSQNDIIVWKYPSLTQVAKL
TGHSYRVILAMSPDEALVTGADETLRFWMVFSKTRSTKESVSLNLFTRIR"

BASE COUNT 939 a 1492 c 1680 g 897 t
ORIGIN

alignment_scores:
Quality: 1294.00 Length: 495
Ratio: 3.469 Gaps: 14
Percent Similarity: 75.354 Percent Identity: 54.343

alignment_block:
US-09-701-572-2 x AB033068 ..

Align seg 1/1 to: AB033068 from: 1 to: 5008

6 AsnArgAsnProProTherTherValArgAspAsnSerProPro 22
|||||
83 AATGAGAACACGATGCGACGCGATGCGCGACCGCTGACGCC 132
22 ocLupProSerProGluSerLeuArgHisValSerArgMetLeuSera 39
|
133 TCCCAAGCTCCCAAGTCTCC..... 151
39 snHisTyrTherSerProSerArgThiLeTyrSerAspArgPheIlePro 55
|||||
152TCGCCACGACAG.....CAGGAGACCGCTTCATCC 184
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
|||||
185 TCCAGAGCCGAGCCAACTGAGCGGAACTTCCACAGGATTAACGAGAA 234
69ThrProThrGluGlyArg.....AspAspSerSer..... 79
235 TGAAGAGTCTCCCATGACACCGGAAAGCCAAAGGCGCCACCTCAAGACA 284
80AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
|||||

285 ACGGCAAAAGACGCGCTGCTACTCTGCCCTGCTCAAGATGAGCTGCTG 334
91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
335 GGGCCCGCATGAGAGAGTGCAGAGACCCGACACTGAGAGCGGAGGCT 384
104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyrLysThrG 121
385 GCAGCCCTC...ACGCCGAGAAAGAGGCTGTTCACGATCTCCCTTA 431
121 LuThrArgLysSerMet.....HisSer 128
432 GCACCAAGGCTCCAGCCCGATGAGCGCAAGAGTGTCTCCCTACTTC 481
129 LeuSerProPheMetAspAspAspPheValProGlyValAsnHisSerP 145
482 CTGTCTCCGCTGACCAAGAGAC.....CAGAACCTCTCCGCTCC 525
145 oValLysAlaProArgLysValProArgSerProTyrLysValLeuAspA 162
526 CCGGAAGCCACCCGCAAGATCTCCAAGATCCCTTCAAGGTGCTGAGCG 575
162 LaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTrpSerSer 178
576 CGCCCGAGCTGCAGAGACGACTTCTACTCTCAATGTGTGAGACTGTCTCC 625
179 HisAsnValLeuAlaValGlyLeuGlyAsnValTyrLeuTrpAsnAl 195
626 CTCAAATGCTGACAGCTGGGGCTGAGGACCTGCTGCTACTCTGAGAGTGC 675
195 acYSerSerLysValThrLysLeuCysAspLeuGlyValAsp...AspC 211
676 CTGACACAGCAGAGTGACGGGCTGTGTGACTGTGACTGTGAGGAGGACT 725
211 yValCysSerValGlyTyrAlaGlnArgGlyThrHisLeuAlaValGly 227
726 CAATGACCTCTGCGGCTGTGTGAGCGGGGAACTGTGGGGGGTGGG 775
228 ThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysLysLys 244
776 ACACACAGGAGCTTCTGTCAGATCTGGAGCGCAAGCCGCAAGGAAGCT 825
244 eaYSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerS 261
826 GTCCATGTGTGAGAGGCCACACGCGACCGCTCGGGCGCTGCTGGAATG 875
261 eSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGlnArgAsp 277
876 CTGAGCAGCTGTCTGCTCCGGAGCGCGACCGCATGATCTGCAAGAGGAC 925
278 IleArgThr.....GlnGluAspPheValSerLysLeuSerGlyH 291
926 ATCCGACCCCGCCACATGCACTGCGAG.....CGGGGCTGCAAGGCCA 969
291 sLysSerGlnValCysGlyLeuLysTrpSerTyrAspAsnArgGluLeuA 308
970 CCGGACAGAGGTGTGCGGCTCAAGGTGTCACAGACACACACTCTCTCG 1019
308 laSerGlyGlyAsnAspAsnLysLeuPheValTrpAsnGlnHisSerThr 324
1020 CTTGGGGGGGACGACACAAAGCTGTGTGCTGGAATCACTGAGGCTG 1069
325 GlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleAl 341
1070 AGCCCGCTGACGAGTACACGAGACCACTGCGCGCGCTGAAGGCCATCGC 1119
341 aTrpSerProHisLeuHisGlyLeuLeuAlaSerGlyGlyGlyThrAla 358
1120 CTGCTCCCAATCAGACGAGGCTGTGCTGCTCGGGCGGCGCACACTG 1169
358 spaArgCysIleArgPheTrpAsnThrThrTrpAsnSerHisLeuSerCys 374
1170 ACCGCTGATCTCGCTTGTGAACAGGCTGACAGGACAAACCACTGCACTGT 1219

```

375 MetAspThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValas 391
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1220 ATGACACAGGGGCTCCCAAGGTGCTATGCGCTGCTCAAGACGACCA 1269
391 ngLueValSerThrHisGlyTyrSerGlnAsnGlnIleValTrpA 408
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1270 CGAGCTGCTGAGCAGCAGCGCTACTACAGAACCATCTGCTGCGA 1319
408 rGlyTrpProMetSerLysLeuValaThrLeuThrGlyHisThrTrp 424
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1320 AGTACCCCTCCCTGACGACGAGTGGCCAGCTGACGCGATCTAC 1369
425 ValLeuTyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAl 441
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1370 GTGCTGTACTGACATGTCCTGATGGGAGGCGATGTCACGTGCTC 1419
441 agLysAspGlyThrLeuArgPheTrpAsnValPheProSerProLysSerG 458
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1420 TGGAGACGAGACCTGAGGTCTGAGAACGCTTT.....AGCAAAACCC 1463
458 InAsnThrGlySerGlnIleGlyAlaLeuSerLeu 469
      ::::::::::::::::::::::::::::::::::::::::::::::::::::
1464 GTTCGACAAAGAGAGTGTGTCTGTCTGCTCAACCTC 1498
seq_name: gb_AB013462
seq_documentation_block:
LOCUS AB013462 1491 bp mRNA linear PRI 23-NOV-1999
DEFINITION Homo sapiens mRNA for Fzrl, complete cds.
ACCESSION AB013462
VERSION AB013462.1 GI:6463678
KEYWORDS Fzrl; Fzrl.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1491)
Kotani, S., Oyama, T. and Todokoro, K.
Human homologue of Filizy-related protein
Published Only in Database (1999) In press
2 (bases 1 to 1491)
Kotani, S., Oyama, T. and Todokoro, K.
Direct Submission
Submitted (29-APR-1998) Kazuo Todokoro, The Institute of Physical
and Chemical Research (RIKEN), Molecular Cell Science Laboratory;
3-1-1, Koyada, Tsukuba, Ibaraki 305-0074, Japan
(E-mail: todokoro@rcc.riken.go.jp, Tel: 81-298-36-9075,
Fax: 81-298-36-9090)
FEATURES
Source
1..1491
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1491
/gene="Fzrl"
1..1491
/gene="Fzrl"
/note="Fizly-related protein"
/codon_start=1
/product="Fzrl"
/db_xref="GI:6463678"
/translation="MDQYERLLNQIVIONENTMRYTEMRTLLTPASSPVSPSKH
GRFVIPSRAQAMSVNFRINENKSPSONRKAQKATGNGDGLAYSLKNEILGA
GIEKVDQPTEDRLQSPPEKGLFTYSLSTRSPDGNVSPYSLSPVNSOKL
LSPKRPTRKISKIPKVLADDELDFYLNVDMSLNIVSVIGCTVYLSACTSO
VTRICDLSVSGSVTSVSGSERGNLYAVNTGHTGPTQINDAAGKLSMLEGHTAVGA
LAWNAEQLSSGSRDRRIIDRLRTPPLOSERNLQGRDEVCGLKSTHDLIASGND
NKLVWNHSSLSPVQOYTEHLAVALAIAVSPHOGILASGGGTADRCIRFWNTLGGP
LOCIDTGSQVNCMLAMSKANELVSTHGYSQNDILVMKYPSILOVAVAKLGHSHRYLYLA
MSPDEAIIYTGAGDETLRFWNVFSKTRSKYKWSVSLNLTFRIR"
BASE COUNT 312 a 479 c 452 g 248 t
ORIGIN

```

```

alignment_scores:
  Quality: 1286.50      Length: 496
  Ratio: 3.458
  Percent Similarity: 75.000      Percent Identity: 54.234
alignment_block:
US-09-701-572-2 x AB013462 ..
Align seg 1/1 to: AB013462 from: 1 to: 1491
6 AsnArgSerProProThrSerThrValArgAspAsnSerProPro 22
  |||
49 AATGAGAACAGATGCCAGCGGTGACAGATGCGGGGACCCGAGCC 98
  |||
22 OGluProSerProGluSerLeuArgHisValSerArgMetIleAsnSera 39
  |||
99 TGCCAGCTCCCAAGTGC..... 117
39 snHISlyThrSerProSerArgThrIleTyrSerAspArgPheIlePro 55
  |||
118 .....TCCGCCAGCAAG.....CAGGAGACCGCTTCATCC 150
56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
  |||
151 TCCAGAGCGCGGAGCCACTGAGCGGTGACTCCACAGATTTACGAGAA 200
69 .....ThrProThrGluGlyArg.....AspAspSerSer..... 79
  |||
201 TGAGAGTCTCCAGTCAGAACCGGAAAGCCAGACGCGCCCTCAGACA 250
80 .....AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
  |||
251 ACGGCAAGAGAGCGCTGCTACTCTGCTCTGCTCAAGATGAGCTGCTG 300
91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
  |||
301 GGTCCCGCATCGAGAGGTGAGAGACCGCAGACAGTAGAGACCGAGGCT 350
104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyrLysThrG 121
  |||
351 GCAGCCCTCC...ACGCTGAGAAAGAGGTGTTGACATATCCCTTA 397
121 IuThrArgGlnSerMet.....HisSer 128
  |||
398 GCACCAAGCGCTCCAGCCCGATGAGCGCAAGATGTCCTCCCTACTCC 447
129 LeuSerProPheMetLaspAspAspPheValProGlyValAsnHisSerPr 145
  |||
448 CTGCTCTCCGCTCAGCAACAGAGC.....CAGAACCTGCTCCGCTCC 491
145 oValLysAlaProArgLysValProArgSerProTyrLysValLeuAspA 162
  |||
492 CCGGAAACCCACCCGCAAGATCTCCAGATCCCTTCAAGGCTGAGAG 541
162 IaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTrpSerSer 178
  |||
542 CGCGCGAGCTGCGAGAGCACTTACTCTCAATGTGTGAGACGTGCTGCC 591
179 HisAsnValLeuAlaValGlyLeuGlyAsnCysValTyrLeuTrpAsnAl 195
  |||
592 CTCAAATGCTGCTAGCTGGGGCTAGGACACCTGCTACTCTGAGAGTCC 641
195 acySerSerLysValThrLysLeuGlyAspLeuGlyValAsp...AspC 211
  |||
642 CTGTACACAGCAGATGACCGGCTGTGACCTCTCAGTGAAGGAGGACT 691
211 yValLysSerValGlyTrpAlaGlnArgGlyThrHisLeuAlaValGly 227
  |||
692 CAGTGACCTCCGCGGCTGCTGACGCGGGGAGACCTGTGTGCGGTGGCC 741
228 ThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysLysIle 244

```



```

39 snHsIytrHsrProSerArgThrIleTyrSerAspArgPheIlePro 55
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
341 .....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
56 SerArgSerAlaSerIlePheAlaLeu.....PheAspIleAsn..... 68
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
374 TCGCGGGCGGGCCAACTGAGCGTGAATCTCCACAGATCATGATAA 423
69 .....ThrProThrGluGlyArg.....AspSerSerSer..... 79
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
424 TTAGAAGTCCGCCACAGACCAAGCCAAAGAGCCACCTCGGACA 473
80 .....AlaTyrThrThrIleLeuArgThrAlaLeuPhe 90
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
474 ATGGCAAGAGCGGCTGGCTACTCCGCACTGCTGAGATGAGTGTG 523
91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
524 GGTGGCGGCATCGAGAGGTTCCAGACCCAGACGGAGAGCGGCGCT 573
104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyrThrG 121
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
574 GGAGCGGTCC...ACACAGAGCAGACAGGGGCTCTTACGATTCCTCA 620
121 LuThrArgGlnSerMet.....HisSer 128
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
621 GGAGCAGAGCGCTCGAGTCCAGATGATGCAATGACGTGCCCATACT 670
129 LeuSerProPheMetAspAspAspPheValProGlyValAsnHisSer 145
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
671 CTGTCCCGCTTACGACAAAGT.....CAGAGCTGTGCGGTCCAC 714
145 oValLysAlaProArgLysValProArgSerProTyrLysValLeuAsp 162
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
715 ACGGAGCCACACGCAAGATCTTAAGATTCCTCAAGGTGCTGGAGC 764
162 IaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTyrSer 178
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
765 GCCCAGAGCTTCAGAGCAGCTCTACCTCAACTGTGGTGGTCTGCC 814
179 HisAsnValLeuAlaValAlaGlyLysCysValTyrLeuTyrPasn 195
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
815 CTCACAGGCTCTACGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 864
195 acYSerSerLysValThrLysLysCysAspLeuGlyValAsp...Asp 211
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
865 ATGCACACAGCGAGTACCGGCTGTGACCTCTCTTAAAGGGGACT 914
211 ysaValCysSerValGlyTyrPalaGlnArgGlyThrHisLeuAlaVal 227
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
915 CAGTACATTCGGTGGCTGTGCTGAGCGGGGAACCTGTGCGAGTAG 964
228 ThrAsnAsnGlyLysValAlaGlnIleTyrPasnAlaArgCysLys 244
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
965 ACACACAAAGGCTCTGCAATCTGGGAGCTGCTGCTGGGAGAACT 1014
244 earGserMetGluGlyHisArgLeuArgValAlaLeuAlaTyrPasn 261
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1015 GTCCATCTGTGAGAGGCCACACAGCAGAGTGGGGCGCTGGCATG 1064
261 earSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGlnArg 277
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1065 CTGACACAGTGTCACTGTGAGCCGTGACCCATGATCTTACACGG 1114
278 IleArgThr.....GlnGluAspPheValSerLysLeuSerGly 291
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1115 ATCCGACACACACCCCTGACGCTCAGAG.....CGGCGGTGCGAG 1158
291 slySerSerGluValCysGlyLeuLysTyrSerTyrAspAsnArgGlu 308
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1159 CCGGAGGAAAGTGTGGCTTAAGTGTGTCCACAGACACACACGCTGT 1208

```

```

308 IaSerGlyGlyAsnAspAsnLysLeuPheValTyrPasnGlnHisSer 324
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1209 CCTGGGGGGGCAATGACACAGCTGCTGCTGGGAACCTAGTCTA 1258
325 GlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAla 341
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1259 AGCCTGTGACAGCAGTATACGAGACACCTGGCAGCGCTGAAGCT 1308
341 aTrpSerProHisLeuHisGlyLeuLeuAlaSerGlyGlyTyrPala 358
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1309 CTGGTCCCAACACACGATGATGCTGCTGCTGCTGCTGCTGCTG 1358
358 spArgCysIleArgPheTyrPasnThrThrTrasnSerHisLeuSer 374
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1359 ACGCTGTATCCGATCTGGAACACCTCTGACAGTCAAGCAGCAGT 1408
375 MetAspThrLysSerGlnValCysAsnLeuValTyrSerLysAsn 391
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1409 ATGACACAGAGCTCACAAAGTGTGCAACCTGGCTGCTGCAAGC 1458
391 nGluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleLeuVal 408
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1459 TGAAGTGTGAGACACATGCTACTCTACAGAACAGATCTGTGTGA 1508
408 rGlyProThrMetSerLysLeuAlaThrLeuThrGlnHisThrTyr 424
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1509 AGTACCGCTCCCTTACGACAGTGGCCAACTGCTGCACTATATCG 1558
425 ValLeuTyrLeuAlaIleSerProAspGlyGlnThrIleValThr 441
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1559 GTCTCTACTCTGCGCAATGCTCCCTGATGGGAGCCATGATCACC 1608
441 agLysAspGluThrLeuArgPheThrPasnValPheProSerProL 458
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1609 TGGAGATGACACCTGAGGTTCTGGAATGTCTTC.....AGCAA 1652
458 InAsnThrGlnSerGluIleGlyAlaLeuSerLeu 469
   |||||.....TCACCCAGCAG.....CATGGTACGGCTTCATCCCC 373
1653 GCTCTACAAAGAAATCTGTGTCTGTCTCAACTC 1687

seq_name: gb_ro:AF083809
seq_documentation_block: 2258 bp mRNA linear ROD 01-SEP-1999
LOCUS AF083809
DEFINITION Mus musculus fizzy-related protein (Fyr) mRNA, complete cds.
ACCESSION AF083809
VERSION AF083809.1 GI:5813824
KEYWORDS
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2258)
AUTHORS Jin,D.-Y. and Jeang,K.-T.
JOURNAL Characterization of mouse fizzy-related protein
REFERENCE 2 (bases 1 to 2258)
AUTHORS Jin,D.-Y.
JOURNAL Unpublished
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1998) LHM, NIAID/NIH, 9000 Rockville Pike,
Bethesda, MD 20892-0460, USA
FEATURES
source
1..2258
/organism="Mus musculus"
/db_xref="taxon:10090"
1..2258
/gene="Fyr"
143..1624
/gene="Fyr"
/function="cell cycle regulator"
/codon_start=1
/product="fizzy-related protein"
/protein_id="AAD52029.1"

```


seq documentation block:

LOCUS XLRNAFRP 2649 bp mRNA linear VRT 08-SEP-1997
 DEFINITION Xenopus laevis mRNA for fizzy-related protein.
 ACCESSION Y14163
 VERSION Y14163.1 GI:2326942
 KEYWORDS fizzy-related protein; far gene.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodinae; Xenopus.

REFERENCE
 1 (bases 1 to 2649)
 Slight, S.J. and Lehner, C.F.
 Drosophila fizzy-related down-regulates mitotic cyclins and is
 required for cell proliferation arrest and entry into endocycles
 Cell 90 (4), 671-681 (1997)

JOURNAL
 MEDLINE
 97433078
 2 (bases 1 to 2649)
 Lehner, C.F.
 Submitted (04-JUL-1997) C.F. Lehner, Department of Genetics,
 University of Bayreuth, 95440 Bayreuth, FRG

FEATURES
 source
 1. 2649
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /issue_type="ovary"
 10. 99
 /gene="fzr"
 100. 1581
 /gene="fzr"
 100. 1581
 /gene="fzr"
 /codon_start=1
 /product="fizzy-related protein"
 /protein_id="CAA4576.1"
 /db_xref="GI:2326943"
 /translation="MDQYERLLRQINLQNTIFPCASEMRRLTPSPMSPSKH
 GDRFISRAGAMSIINFHRIENERSPSQNRKAKATADSGDGLAYSALIKNELGA
 GIEKVDPTQEDRRLDPTPEKSKLTPYLSKSRASPDGNCVSPISPSVNSKOKL
 LNSPKPTKISKIPKVIDAPLQDFYLNLDVSSLVNLSVGLQCYLYMSACTSQ
 VRLCDLSVSGSVTSVWSEKGNLVAAGTHKQVAVDASAGKISTLEGHTARVGA
 LAMNADQLSGSRDMLQIDIRTPVQSEKRIQGRQVCGIKSTDHQLASGND
 NKLWNHSSLSPVQOYTEHLAAVKAIAKSPHOGHLAGSGTADSCIFFWTLIGOP
 LQCIDGSOVCNLAISKHANEIYSTHGYSONOILWKPYSILQVAKLGHYSRVLYLA
 MSPDGRALIVTGADELIRFNNVPSKTRSTKESVIANLPTTRR"
 1582. 2649

3' UTR
 BASE COUNT 758 a 541 c 598 g 752 t
 ORIGIN

alignment_scores:
 Quality: 1285.00 Length: 495
 Ratio: 3.427 Gaps: 14
 Percent Similarity: 75.758 Percent Identity: 53.737

alignment_block:
 US-09-701-572-2 x XLRNAFRP ..

Align seg 1/1 to: XLRNAFRP from: 1 to: 2649

6 AsnATGAsnProProthSerTherValAlrGAspAsnSerProProPr 22
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 148 AATGAGAACACCATATCTTGTCATCAGAAATGAGGAGACATTAAACCC 197
 22 OGluProSerProGluSerLeuAlrGHisValSerArgMetIleAsnSera 39
 | ||||| ||| |||
 198 AACAAACTCTCCCATGCT..... 216
 39 snHISfTyThrSerProSerArgThrlIeTySerAspArgPheIlePro 55
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 217TCACCAAGTAA.....CATGAGACAGGTTTATACCA 249

56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 250 TCCTGGGCGAGGAGCGAATGGACATTACTTCCACAGAAATTAATGACGA 299
 69ThrProThrGluGlyArgAsp.....Asps 77
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 300 TGAATAATCCCGACATCAGACAGAAAGCCAGATGCCAGCGGACAGA 349
 77 erSerSer.....AlaTyThrThrLeuLeuArgThrAlaLeuPhe 90
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 350 GTGGCAAAAGATGGCTTACCTTATTCAGCTCTGCTTAAGATAGAGCTGTG 399
 91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 400 GGAGCTGGTATTGAGAAAGTTCAAGATCCACAGACCTGAAGATAGAGACT 449
 104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyLysThrG 121
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 450 ACAACCATCA...ACGCCAGAAAAAAGAGCCTGTACGTACTCTCTAA 496
 121 IuThrArgGln.....SerMetHisSer 128
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 497 GTTCTTAAGCTGCAGCCAGATGATGAATGAACTGCTCTTAATTCCT 546
 129 LeuSerProPheMetAspAspAspPheValProGlyValAlaHisSerPr 145
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 547 TTGTCACTCTTCAGTTAATAAAG.....CAGAACCTTTGGCGGTCTC 590
 145 ovalLysAlaProArgLysValProArgSerProTyLysValLeuAspA 162
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 591 ACGAAAGCCCTACAGAGAGATTCGAAAGATCCCTTTAAAGTCTCGAG 640
 162 IabProAlaLeuGlnAspAspPheTyLeuAsnLeuValAspTrpSerSer 178
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 641 CACTGAGTTACAAAGATGACTTTTACTCTACCTAGCTAGTGGTATCTC 690
 691 CTCATGTTTAAAGTGGGCTTGGAGACTTGGTACTTATTAATGAGTGC 740
 195 acySerSerLysValThrLysLeuGlyAspLeuGlyValAsp...AspC 211
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 741 ATGTACTAGCCAGGTACTAGTATGTAATCTCTGTGGAGGGGATTT 790
 211 ysalLysSerValGlyTrpAlaGlnArgGlyThrHisLeuAlaValGly 227
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 791 CTGTACTTACTAGTTGGTGTGTCAGACCGGGAACCTGTAGCTGGGG 840
 228 ThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysLysLysI 244
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 841 ACACACAAAGCTTGTGTCAGATTTGGATGATCAGACGAGGAACACT 890
 244 earGSerMetGluGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerS 261
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 891 GTCCACTCTGGAGGACACACACCCCGTGGGGAGCTTGGCTGGAAAG 940
 261 erSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyGlnArgAsp 277
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 941 CAGACCAAGCTTCTCTCGGAGTAGGAGACAGATGATCTTACAAAGAGAT 990
 278 IleArgThr.....GlnGluAspPheValSerLysLeuSerGlyH 291
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 991 ATTGCGACACACCTGTTTCAGTCAGG.....AGAAGCTGTCAAGGCCA 1034
 291 slySerGluValCysGlyLeuLysTrpSerTyraAspAsnArgGluLeuA 308
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 1035 CAGGCAAGAGATGTGGCTCAAGTGTCTCACAGACCCACAGCTCTCTGG 1084
 308 IAserGlyLysAsnAspAsnLysLeuPheValTrpAsnGlnHisSerThr 324
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 1085 CTTCAGAGGAGGAATGACAAACAGCTCTGTGGTGAACCATTTCTACCTG 1134
 325 GlnProValLeuLysTyrcysGlnHisThrAlaAlaValLysAlaIleAl 341

```
1135 AGTCCGTTCACAGTATATCTAGCTCAGTGAAGGCAATTGC 1184
341 aTTPSerProHisLeuHisGlyLeuAlaSerGlyGlyThrAla 358
1185 CTGGTCCCTCATCAACATGAGCTTTGGCTTCAGCGGGGCACTGCTG 1234
358 sPArgCysIleArgPheThrPasnThrThrAsnSerHisLeuSerCys 374
1235 ACCGCTGCATCCGCTTGTGAACACACTGACAGGCCAGCCTTTACAGTGT 1284
375 MetAspThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValAs 391
1285 ATTGATTAAGTGGCTCTCAAGTCTGTAACCTGGCCGTGCTAAGCATGCAGAA 1334
391 nGluLeuValSerThrHisGlyTySerGlnAsnGlnIleIleValTrpA 408
1335 TGAACGTGTAAGCACACATGCTACTCTCAAAATCAGATATTAGTATGGA 1384
408 rGlyTrpProThrMetSerLysLeuAlaThrLeuThrGlyHisThrTyArg 424
1385 AATATCCATCGCTTACTACAGGTGCGCAAGTTGACTGGACATTCTACCGT 1434
425 ValLeuTyrlleuAlaIleSerProAspGlyGlnThrIleValThrGlyAl 441
1435 GTTCTCTACCTGGCTATGTACACAGATGGGAGGCCATTGTTACAGGAGC 1484
441 aGlyAspGluThrLeuArgPheThrPasnValPheProSerProLysSerG 458
1485 AGGTGATGAACGCTGAGATCTTGGAATGTTTTT....AGCAAAACAC 1528
458 InAsnThrGluSerGluIleGlyAlaLeuSerLeu 469
1529 GCTCCACAAGAGAGTGTGTCTGGTCTPAAACCTG 1563
```


alignment block:
US-09-701-572-2 x AA237024 ..

Align seg 1/1 to: AA237024 from: 1 to: 2006

```

1 MetAspGlyThrGlyAsnArgAsnProProThrSerThrValArgAs 17
182 ATGGACGGAAACCGGTAATCGAATCCACACCGACTCCACCGTCAGACA 231
17 pAsnSerProProGluProGluProGluProGluProGluProGlu 34
232 CAAATTCCTCCACCGCCCTGAGACCACTACCGGAGAGTCTCCGATGTAAGCC 281
34 rMetLeuSerSerAsnHisThrThrSerProSerArgThrLeuSer 50
282 GATATGATCAACAGCAACCAATACCTCCTTCGAAACAATCTACTGC 331
51 AsparGlyPheLeuProSerArgSerAlaSerLysPheAlaLeuPheAsp 67
332 GATAGGTTGATTCGAGTAGATCTGCTGGAATTCGCTTGTTGATAT 381
67 eAsnThrProThrGluGlyArgAspAspSerSerSerAlaThrThrL 84
382 CAATACCTCGACAGAGAGCGGATGATGTTCCAGCGCTTATACGACTC 431
84 euLeuArgThrAlaLeuPheGlyProAspValAlaGlyProValThrPro 100
432 TTCTGGAACCGCGCTTGTGGACCGGATGTTCCGCTCCGCTTACGCG 481
101 GluLysThrAspSerProSerMetThrLeuProAsnArgAsnLeuPheAr 117
482 GAAAAAACCGACTCCCGCTGATGACATGCGCAATAGCAATATTTTAA 531
117 gThrLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMet 134
532 GATATAGAGGAGACGACAGACATCCATCGACATCGCTTCCGCTTAA 581
134 sPAspAspPheValProGlyValAsnHisSerProValLysAlaProArg 150
582 ATGATGATTTGTTCTCTGGTGTATCATAGTCGGTTAAGCTCTTACG 631
151 LysValProArgSerProThrLysValLeuAspAlaProAlaLeuGlnAs 167
632 AAGGTTCCCGATCCCTTATAGGTTTGGATGCACCTGCTTGCACAGA 681
167 pAspPheThrLeuAsnLeuValAspTrpSerSerHisAsnValLeuAla 184
682 TCATTTTATCTGAATCTGATGATGATGCTTCACACATGATGGGCTG 731
184 aGlyLeuGlyAsnCysValThrLeuTrpAsnAlaCysSerSerLysVal 200
732 TTGGTTGGGTAACGTGCTATCTCTGGAATGCTTGTAGCAGCAAGTA 781
201 ThrLysLeuCysAspLeuGlyValAspAspCysValCysSerValGlyTr 217
782 ACTAATATATGATTTGGGGGTGATGATGTTGTTCTGTTGGTGG 831
217 pAlaGlnArgGlyThrHisLeuAlaValAlaGlyThrAsnAsnGlyLysVal 234
832 GGCTAAAGCTGGTACTCATCTTGGCTGTGGAACATACATGATGAAGTTC 881
234 InLeuTrpAspAlaAlaArgCysLysLysLysLeuSerMetGluGlyHis 250
882 AGATTGGGATGACAGCAAGATGCAAGATGAATCAATGAGAGGCCAT 931
251 ArgLeuArgValAlaLeuAlaTrpSerSerSerLeuLeuSerSerG 267
932 CGGTACCTGTCGGGGCTTGGCCGAGTCACTCTTTGTTGCTTCCG 981
267 gGlyArgAspLysAsnLeuTrpGlnArgAspLeuArgThrGlnLysP 284
982 TGGACGGATAGAAATATTTATCAACGAGATATACGACACAAAGATT 1031

```

```

284 heValSerLysLeuSerGlyHisLysSerGluValCysGlyLeuLysTrp 300
1032 TTGTTAGTAAACTGTCAGACAAATCAGAGGTTTGTGAGCTGAACTGG 1081
301 SerTrpAspAsnArgGluLeuAlaSerGlyGlyAsnAspAsnLysLeuP 317
1082 TCATATGATTAACCTGAGTTGGCATCTGGAGGAAATACAACAATTTGTT 1131
317 eValTrpAsnGlnHisSerThrGlnProValLeuLysTrpCysGluHis 334
1132 TGTGGATCAACATCAACCCAGCTGCTCCCAATAGTACTGAGACACA 1181
334 hAlaAlaValLysAlaAlaLeuTrpSerProHisLeuHisGlyLeuLeu 350
1182 CAGCAGCTTTAAAGCTTATGATGATGCTCTCTCATCTTCATGAGACTTCT 1231
351 AlaSerGlyGlyGlyThrAlaAspArgCysAlaLeuArgPheTrpAsnThr 367
1232 GCATCTGGAGAGGAGACTGCAGATAGATGATTCGTTTGGAAATACAC 1281
367 rThrAsnSerHisLeuSerCysMetAspThrGlySerGluValCysAsn 384
1282 CACAACTCACACCTTACTGATGAGACACTGGAGACAGTTTGGCAATC 1331
384 euValTrpSerLysAsnValAsnGluLeuValSerThrHisGlyTrpSer 400
1332 TTGCTGCTGCTCAAAATGTCACAGCACTAGTAAGCACACATGGGTTACT 1381
401 GlnAsnGlnLeuLeuValTrpArgTrpProThrMetSerLysLeuAla 417
1382 CAGAACCAAGATTTATGTTGGAGATACCCCATATGTCAAAGCTGGCGAC 1431
417 rLeuThrGlyHisThrTrpArgValLeuTrpLeuAlaLeuSerProAsp 434
1432 TCTTACCGGCATCTTATAGGTTCTTATCTTATGCTCCAGATG 1481
434 LysGlnThrLeuValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsn 450
1482 GACGACATATTGTAAGCTGAGAGCTGAGATGAAACCGTTAGCTTGGAAT 1531
451 ValPheProSerProLysSerGlnAsnThrGluSerGluLeuGlyAlaLe 467
1532 GTTTCCTCCCTCCCTTAATCAGAAATACGAAAGTGAATCGGACATT 1581
467 uSerLeuGlyArgThrThrLeuArg 475
1582 ATCTCTTGAAGAACTACTATACAG 1606

```

seq_name: /SIDSI/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AACT7458

seq_documentation block:
ID AAC77458 standard: cDNA: 5905 BP.
XX AAC77458;
XX
XX 08-FEB-2001 (first entry)
XX
XX
DE Human ORF3013 polynucleotide sequence SEQ ID NO:6025.

Human; open reading frame; ORF; detection; cytosolic; hepatotropic;
vulnerary; antiparietal; antiparkinsonian; noctropic; neuroprotective;
anticonvulsant; osteopathic; antiaerthetic; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antinflammatory;
antiviral; antibacterial; antifungal; antihemetic; antithyroid;
antianemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antiinflammatory disease; coagulation;


```

325 GlnProValLeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleAl 341
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1297 ACCCCGTCGACAGTACACGAGCAGCAGTGGGCGCTGGAAGGCCATGCG 1346
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
341 arPserProHisLeuHisGlyLeuLeuAlaSerGlyGlyThrAla 358
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1347 CTGGTCCCACTCAGCACAGGCTCTGCGCTCGGGGGGGCGGACAGCTG 1396
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
358 sPaRgCysIleArgPheTrpAsnThrThrAsnSerHisLeuSerCys 374
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1397 ACCGCTGATCGCTTCTGACACAGCTGACGACGACAACTGCAAGTGT 1446
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
375 MetAspThrGlySerGlnValCysAsnLeuValTrpSerLysAsnValAs 391
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1447 ATCGACACGGCTCCCAAGTGTGCAATGTGCTGCTCCAGCAGCCAA 1496
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
391 nGluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleIleValTrpA 408
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1497 CGAGCTGGTGAACAGCAGCGCTACTCAGACAGACAGATCCTGTCTGGA 1546
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
408 rGlyTrpThrMetSerLysLeuAlaThrLeuThrGlyHisThrTyrArg 424
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1547 AGTACCCCTCCTGACCCAGGTGCGCAAGCTGACCGGCGACCTCCTAC 1596
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
425 ValLeuTyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAl 441
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1597 GTGCTGTACTGTGCAATGTCCCTGATGGGGGCCATGCTGACTGTGTC 1646
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
441 aGlyAspGluThrLeuArgPheTrpAsnValPheProSerProLysSerg 458
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1647 TGGAGACGAGACCTGAGGTCTTGACAGCTTTAGCAAAACCCCTTGA 1696
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
458 IAsnThrGluSerGlu...IleGlyAlaLeuSerLeu 469
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
1697 CAAGGTAAAGTGGAGTGTGTGTGCTCAACCTC 1734
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABU05493
seq_documentation_block:
ID   ABU05493 standard; cDNA; 2645 bp.
XX
AC   ABU05493;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 10961.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PE   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
XX
PT   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW.
XX
DR   WPI; 2001-656860/75.
XX
DR   P-PSDB; ABB61390.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX

```

```

PS   Claim 1: SEQ ID NO 10961; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent.
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC   sequences (ABU01840-ABU16175) and the encoded proteins
CC   (ABU57737-ABU72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2645 bp; 766 A; 630 C; 619 G; 630 T; 0 other:

alignment_scores:
      Quality: 1276.00      Length: 474
      Ratio: 3.535          Gaps: 11
      Percent Similarity: 76.160      Percent Identity: 54.852

alignment_block:
US-09-701-572-2 x ABU05493 ..

Align seg 1/1 to: ABU05493 from: 1 to: 2645

22 ProGluProSerProGluSerLeuArgHisValSerArgMetIleAsn 38
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
507 CCCGAGTACAGAGAGCCAGCTGGAAGCACTGACAGCTCTGTGCGAGAA 556
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
38 r.....AsnHisTyrThrSerProSerArgThrIleTyrSerAspArg 53
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
557 TGTGTTCACAACTTCGAGTGTGTCACACTAGCCCACTCTCTGACCGCT 606
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
53 heIleProSerArgSer.....AlaSerLysPheAlaLeuPhe 65
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
607 TCATACCCCTGACAGCGCTACACAACTGCGACGAACTTTGGG..... 650
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
66 AspIleAsn.....ThrProThrGluGlyArg 74
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
651 TCATATCAACAACTGCAATGACAACTGCCCGCAGACAGTAGAAGCAGCG 700
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
74 gAsp.....AspSerSerAlaTyrThrThrLeuLeuA 86
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
701 GGACTGGGGGGAACGCGACGCAATAGTCTGCTACTCTGCTCTGCA 750
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
86 rGlnrAlaLeuPheGlyProAspVal.....AlaGlyPro 97
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
751 AGAAGCAGCTCTCTGCGCAATCGACGACGAGTGAAGACCGCGCGAG 800
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
98 ValThrProGluLysThrAspSerProSerMetThrLeuProAsnArgAs 114
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
801 GAGCGGAATGAGAAATGCCCTACACCGCGCC.....GCAAGCGGAG 841
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
114 nIlePheArgTyrLysThrGluThrArgGlnSerMet..... 126
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
842 TCTCTTCAAGTACAGTACACCACCAAGACGACTACATAGCGAGTGTG 891
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
127 ..HisSerLeuSerProPheMetAspAspAspPheValProGlyValAsn 142
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
892 CGTACTGCTGTGTACCCGTCAGCGCCAAAGT.....CAGAAGCTGTG 935
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
143 HisSerProValLysAlaProArgLysValPheArgSerProTyrLysVa 159
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
936 CGATGCCCGCGCAAGGCTACGCGCAAAATCTCTGCGATTCCTTCAAGT 985
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
159 IleAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspT 176
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
986 GCTAGACGCGCGCGAGTTCGACGACGACTTCTATCTGAACCTGTGAGCT 1035
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::
176 rpSerHisAsnValLeuAlaValGlyLeuGlnAsnValTyrLeu 192
      :::::::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

1036 GGTCTGCGACAAAGCTACTGCTGAGCCCTGGCCAGCTGTCTACTG 1085
193 TTPAANAIACTSSerSerlyValThrlyLeuYsaSplLeuGlyValAs 209
1086 TGGACGGGTGCTACCACTGACGCTTACCGCTGATGATCTCAGTCCGGA 1135
209 P...AspCysValCysSerValGlyTTPAlaGlnArgGlyThrHisLeuA 225
1136 TCCGAATACGCTGACCTGCTGTGTGAGAACGACGCTGGCAACACCGTGG 1185
225 IValGlyThrAsnAsnGlyLyValGlnIleTTPAspAlaAlaArgCys 241
1186 CCGTGGGACACATCAGCTACCTGACCCCTGTGGAGTGTGGCGCGCAAT 1235
242 LysLysIleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAl 258
1236 AAGCAGATCAACAATGATGATGCGCTGCGCGCTGTGGCGCTTGGC 1285
258 ATPSerSerSerLeuLeuSerSerGlyLyArgAspLysAsnIleTyrG 275
1286 ATGAAACAGTGAATCTGTGAGCGGGTGGCGAAGCGTGTGATCATAC 1335
275 IValAspIleArgThr...GlnGluAspPheValSerLyLeuSerGly 290
1336 AGCGGATACGAAAGCGCCCACTGCAATCGAGCGCAGATTTGGCCGGA 1385
291 HisLysSerGluValCysGlyLeuLysTTPSerTyrAspAsnArgGluLe 307
1386 CATCGGCGAGAGGTGTGCGACTGAAATGCTACCGGATTAATCAATCTT 1435
307 uAlaSerGlyGlyAsnAspAsnLysLeuPheValTTPAsnGlnHisSerT 324
1436 GCCCAGTGGCGGCACGATTAATCGCTGTATGTGTGAAATCAGCATTCG 1485
324 hGlnProValLeuLysTyrCysGluHisThrAlaAlaValLysAlaLe 340
1486 TGAATCCCGTACAAATCATACAGGAGCATATGCGCGCTGTAAGCGCATC 1535
341 AlaTTPSerProHisLeuHisGlyLeuLeuAlaSerGlyGlyTThAl 357
1536 GCGTGTGCGCCCATCACCACGAGTCTCTGGCCACGCGCGGTGGACAGC 1585
357 AsnArgCysIleArgPheTTPAsnThrThrAsnSerHisLeuSerC 374
1586 GGATGAGGTATTCGCTTCTGGAATACCTGACGGGCGCAGCCAGTCAGT 1635
374 yMetAspThrGlySerGlnValCysAsnLeuValTTPSerLysAsnVal 390
1636 GCGTGGACAGGGGCTGCGAGGTTTGCATCTGCGCTGCTCCAAAGCATCC 1685
391 AsnGluLeuValSerThrHisGlyTyrSerGlnAsnGlnIleIleValTr 407
1686 TCGGAGCTGTGCTCCACGACGCTACTCGCAACACGATACCTGTGTG 1735
407 pArgTyrProThiMetSerLysLeuAlaThrLeuThrGlyHisThrTyrA 424
1736 GAAATATCCCTCCCTGACGCAAGTGGCCAAAGCTGACGGGCATTCGTATC 1785
424 rGValLeuThrLeuAlaIleSerProAspGlyGlnThrIleValThrGly 440
1786 GTGTGCTCTATCTGGCGGTGAGTCCCGATGTGTGAGCTATTGTACGGGC 1835
441 AlaGlyAspGluThrLeuArgPheTTPAsnValPheProSerProLysSe 457
1836 GCCGGGAGAGAGCGTGGGATTGTGAAACGATTCACAGAGCGCGCAG 1885
457 rGlnAsnThrGluSerGluIle 464
1886 TCAGAAAGAGAAACAGTCCGCT 1907
seq_name: /SISL1/gcdata/hold-geneseq/gene-seq-n-emb1/NA2001B.DAT.AAS77749
seq_documentation_block:

```

```

ID AAS77749 standard; cDNA; 3940 BP.
XX
AC AAS77749;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #13553.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI MPI: 2001-639362/73.
XX
DR P-PSDB; ABG13562.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX
PS Claim 1: SEQ ID No 13553; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPRO
XX CC at ftp.wipro.int/pub/published_pct_sequences.
XX
SQ Sequence 3940 BP; 799 A; 1246 C; 1221 G; 674 T; 0 other;
XX
alignment_scores:
XX Quality: 1231.50 Length: 588
XX Ratio: 3.293 Gaps: 17
XX Percent Similarity: 63.605 Percent Identity: 46.259
XX
alignment_block:
XX US-09-701-572-2 x AAS77749 ..
XX
Align seg 1/1 to: AAS77749 from: 1 to: 3940
XX
6 AsnArgAsnProProThrSerThrValArgAspAsnSerProProTr 22
||||:||||| ||| |||

```

1423 AATGAGAACAGATGCCACGCTCACAGAGATCGCGGAGACCTTGACGCC 1472
 22 oGluProSerProGluSerLeuArgHisValSerArgMetIleAsnSera 39
 1473 TGGCAGCTCCCGAGGTGCC..... 1491
 39 snHisThrSerProSerArgThrIleTyrSerAspArgPheIlePro 55
 1492TCGCCAGCAAG.....CAGGAGACCGCTTCATCCCC 1524
 56 SerArgSerAlaSerLysPheAlaLeu.....PheAspIleAsn..... 68
 1525 TCCAGAGCGGAGCCCACTGGAGCCGTGAACCTTCACAGGATTAAAGAGAA 1574
 69ThrProThrGluGlyArg.....AspAspSerSer..... 79
 1575 TGGAGAGTCTCCAGTCAGAACCGGAAACCCAGAGACGCCACCTCAGACA 1624
 80AlaTyrThrThrLeuLeuArgThrAlaLeuPhe 90
 1625 ACGGCAAAAGACGGCTGGCCTACTGCTGCTGCTCAAGATGAGGTGCTG 1674
 91 Gly.....ProAspValAlaGlyProValThrProGluLysThrAs 104
 1675 GGTGCGGATCGAGAGAGGTGAGACCCGACACTGAGAGACCGGAGCT 1724
 104 pSerProSerMetThrLeuProAsnArgAsnIlePheArgTyrLysThrG 121
 1725 GAGGCCCTCC...ACCCCTGAGAGAAAGGCTGTGTTCACGTATTCCTTA 1771
 121 LuThrArgLysSerMet.....HisSer 128
 1772 GCACCAAGAGCTCCAGCCCGGATGACGCGCAAGATGTCTCCCTACCTCC 1821
 129 LeuSerProPheMetAspAspAspPheValProGluValAsnHisSerPr 145
 1822 CTGTCTCCGCTCAGCAACAAGAC.....CAGAACTGCTCCGGTCCCC 1865
 145 oValLysAlaProArgLysValProArgSerProTyrLysValLeuAspA 162
 1866 CCGGAAACCCACCGGCAAGATCTCCCAAGATCCCTTCAAGAGTGGCGAG 1915
 162 LeuProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTrpSer 178
 1916 CCGCCGAGCTGAGAGACGACTTCTACCTCAATCTGTGAGCTGCTGCC 1965
 179 HisAsnValLeuAlaValGlyLeuGlyAsnGlyValTyrLeuTrpAsnAl 195
 1966 CTCATGTCTCAGCGTGGGCTAGGACCTGCTGTACTGTGGAGTGC 2015
 195 aCysSerSer..... 198
 2016 CTGTACACGCCAGAAAGGAGACGGATGGAGAGCGGATGGGTAGCAGA 2065
 198 198
 2066 TGGGTGAGCGGATGGGAGAGCGCATGGAGAGCGGATGGAGAGCGGATG 2115
 198 198
 2116 GGAGAGCCGATGGAGAGCGCATGGATGAGCGCATGGAGAGCGCATGGA 2165
 199LysV 200
 2166 CTGGGGCTTCGAGACGAGCGCAAGTCAACACTGACCTTGCCCTCCAGG 2215
 200 aLThrLysLeuGlyAspLeuGlyValAsp...AspCysValCysSerVal 215
 2216 TGACCGGCTCTGTGACTCTCAGTGAAGGGGACTCAGTGAACCTCCGCTG 2265
 216 GlyTrpAlaGlnArgGly..... 221
 2266 GGCTGTCTGAGCGGGGCTCAGAGTGGGCTTGTAGTTCATTTAAGATTATGA 2315

222ThrHisLeu..... 224
 2316 TGAAGACATCTCAGCAGACGCTTGAAGACATATATATGAGGAGGGGAAAC 2365
 225AlaValGlyThrAsnAsnGlyLysValGlnIleTrpAspAla 239
 2366 TGTGTGGCGGTGGCACAACAGGGCTTCGTGAGATCTGGAGGAGCGAGCC 2415
 240 ArgCysLysLysIleArgSerMetGluGlyHisArgLeuArgValGlyAl 256
 2416 GCAGGGAGAGAGCTCCATGTTGGAGGGCCACACGGCAGCGCTGGGGCC 2465
 256 AlaLeuAlaTrpSerSerSerLeuLeuSerSerGlyLysArgAspLysAsn 273
 2466 GCTGGGCTGGAAATGCTGAGACAGCTGTGTGTCGGAGACCGGACCGCATGA 2515
 273 LeTyrGlnArgAspIleArgThr.....GlnIleAspPheValSer 286
 2516 TCTGAGAGAGGAGCATCCGACCCCGCACCTGACAGTGGAG.....CGG 2559
 287 LysLeuSerGlnHisLysSerGluValCysGlyLeuLysTrpSerTyrAs 303
 2560 CGGCTGCAGGGGSCACCGCAGAGAGTGTGGGGCTCAAGTGTCCACAGA 2609
 303 PAsnArgGluLeuAlaSerGlyLysAsnAspAsnLysLeuPheValTrpA 320
 2610 CCACAGCTCTCGCTCGGGGGGCAACGACACACAGCTGTGTGCTGGA 2659
 320 snGlnHisSerThrGlnProValLeuLysTyrCysGlnHisThrAlaAla 336
 2660 ATCAGTCAGGCTGAGCCCGGTGCACAGATACAGGAGCAGCTGGCGGCC 2709
 337 ValLysAlaIleAlaTrpSerProHisLeuHisGlyLeuLeuAlaSerG 353
 2710 GTGAGGCGCATGCTGTGCTCCCAATCAGACAGCGGCTGTGCTGGCGGG 2759
 353 yGlyGlyThrAlaAspArgCysIleArgPheTrpAsnThrThrAsnS 370
 2760 GGGGGGACAGCTGACCGCTGTATCCGCTTGTGAACAGCTCAGAGGAC 2809
 370 eHisLysSerCysMetAspThrGlySerGlnValCysAsnLeuValTrp 386
 2810 AACCACTGCAGGTGTATGACAGCGGGCTCCCAAGTGTGCAATCTGGCTGG 2859
 387 SerLysAsnValAsnGluLeuValSerThrHisGlyTyrSerGlnAsnG 403
 2860 TCCAGACAGCCCAAGAGCTGTGTAGCAGCCAGCGCTACTACAGAGACCA 2909
 403 nIleIleValTrpArgTyrProThrMetSerLysLeuAlaThrLeuThrG 420
 2910 GATCCTGTCTGGAATGACCCCTCCCTGACCCAGGTGGCAAGCTGACCG 2959
 420 LysHisThrTyrArgValLeuTyrLeuAlaIleSerProAspGlyGlnThr 436
 2960 GGCACCTCTACCGGCTGTCTTACCTGCAATGTCCCTGTGGGGAGGCC 3009
 437 IleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnValPhePr 453
 3010 ATGCTCAGCTGTGTGAGAGAGAGACCCCTGAGGTTCTGAGACCTCTTTAG 3059
 453 oSerProLysSerGlnAsnThrGluSerGluIleGlyAlaLeuSerLeuG 470
 3060 CAAGAACCGCTTCGCAAGAGGACGACTTCCTCCGAGGC...CTCGGCTCA 3106
 470 LysArgThrThrIle 474
 3107 GCGCATCTGCCATC 3120
 seq_name: /sidsl/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL05492
 seq_documentation_block:
 ID ABL05492 standard; cDNA; 14283 bp.


```

XX AC ABL05492;
XX XX
XX 26-MAR-2002 (first entry)
XX DT
XX XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ.ID NO 10958.
XX XX
XX KM Drosophila: developmental biology; cell signalling; insecticide;
XX XX
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD
XX 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB61389.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS
XX PS Claim 1; SEQ ID NO 10958; 21pp + Sequence Listing; English.
XX XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57373-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 14283 BP; 4076 A; 2905 C; 2990 G; 4312 T; 0 other;

```

alignment_scores:

```

Quality: 1156.50      Length: 380
Ratio: 3.894          Gaps: 5
Percent Similarity: 78.158      Percent Identity: 59.211

```

alignment_block:

```

US-09-701-572-2 x ABL05492 ..

```

```

Align seg 1/1 to: ABL05492 from: 1 to: 14283

```

```

115 115pheargtyrlysthrgluthtarglnsermethisserieserr 131
    ::::: ::::: ::::: ::::: ::::: :::::
11413 GTCCTTTCGTTGACGACTACATGCGAGTGCCTACTCGTTGCACC 11462
    ::::: ::::: ::::: ::::: ::::: :::::
131 opmetasppaspphevalproglyvalasnhisseriesprovalysa 148
    ::::: ::::: ::::: ::::: ::::: :::::
11463 CGTCAGCGCCCAAAAGT.....CAGAAGCTGTTCGATCGCCGCGAAG 11506
    ::::: ::::: ::::: ::::: ::::: :::::
148 laProArglyValProArgSerProTyrLySValLeuaspAlaProAla 164
    ::::: ::::: ::::: ::::: ::::: :::::
11507 CTAGCGCAAAATCTCTCGCATTCCTTCAAGTGCTAGACGCGCGCGAG 11556
    ::::: ::::: ::::: ::::: ::::: :::::
165 LeuGlnAspSppheTyrLeuAsnLeuValAspTyrSerSerHisAsnVa 181
    ::::: ::::: ::::: ::::: ::::: :::::

```

```

11557 TTGCAGACGACTTCTATCTGACCTGGTCGAGTGTCTGCAGAAAGT 11606
181 115leuValGlyLeuGlyAsnCySValTyrLeuTyrPasnAlaCySers 198
    ::::: ::::: ::::: ::::: ::::: :::::
11607 ACTGCTGTAGGCGCTGGGACACTGTGTCTATCTGTGGAGCGCGTGCACA 11656
198 er..... 198
11657 GTACAGTGAATCTGTATACATTAGCAATATGCTGCACCTTGTGTCAATA 11706
199 .....Lys..ValThrLysLeu 203
11707 AACTAATCAAAATGATTTTCCCATCATGATGATGATGATGATGATGATG 11756
204 CysAspLeuGlyValAsp...AspCySValCySserValGlyTyrAlaG1 219
    ::::: ::::: ::::: ::::: ::::: :::::
11757 TGTGATCTCATGTCGCGATCGCAATATGCGTGCATCGTGTGTGGAACGA 11806
219 nArgGlyThrHisLeuAlaValGlyThrAsnAsnGlyLysValGlnIlet 236
    ::::: ::::: ::::: ::::: ::::: :::::
11807 GCGTGGCAACACCGCGCGCGGCGACACATACGCGTACGTCGACCGTCT 11856
236 rPAspAlaAlaArgCysLysLysIleArgSerMetGluGlyHisArgLeu 252
    ::::: ::::: ::::: ::::: ::::: :::::
11857 GGGATGTGCGCGCCCAATAGACATCAACAACTGAATGCGCATTCGCGCG 11906
253 ArgValGlyAlaLeuAlaTyrPasnSerSerLeuSerSerGlyGlyAr 269
    ::::: ::::: ::::: ::::: ::::: :::::
11907 CGTGTGGCGCGCTTGGCATGAGACATGACATCCGTGTGACGCGGTGCGG 11956
269 gAspLysAsnIleTyrGlnArgPheArgThr...GlnGluAspPheV 285
    ::::: ::::: ::::: ::::: ::::: :::::
11957 AGACCGTTGATCATATACACGCGGATACGACAGACGCGCACTGCAATCGG 12006
285 alSerLysLeuSerGlyHisLysSerGluValCySgLyLeuTyrSer 301
    ::::: ::::: ::::: ::::: ::::: :::::
12007 AGCGGATGTGCGCGCATGCGACATGCGAGGTGTGCGGCTAATAATGTGCA 12056
302 TyrAspAsnArgGlyLeuAlaSerGlyLysAsnAspAsnLysLeuPheVa 318
    ::::: ::::: ::::: ::::: ::::: :::::
12057 CCGATATATCAATACTTGGCGCGATGCGCGCAACGATATCGTTGTATGT 12106
318 11TPAsnGlnHisSerThrGlnProValLeuLysTyrGlySglnHisTrra 335
    ::::: ::::: ::::: ::::: ::::: :::::
12107 GGGATATCAGCATTCCTGGAATCCCGTACATCAATACACGAGCATATAG 12156
335 1aAlaValLysAlaIleAlaTyrSerProHisLeuHisGlyLeuLeuAla 351
    ::::: ::::: ::::: ::::: ::::: :::::
12157 CGGCTGTAAAGCGCATCCGCGTGTGTCGCGCATACACGAGCATCTGGCC 12206
352 SerGlyGlyGlyThrAlaAspArgCysIleArgPheTyrPasnThrTh 368
    ::::: ::::: ::::: ::::: ::::: :::::
12207 AGCGCGGTGGGAACGGCGGATAGTGTATCCGTTCTGGAATACGCTGAC 12256
368 rAsnSerHisLeuSerCySmetAspThrGlySerGlnValCySAsnLeuV 385
    ::::: ::::: ::::: ::::: ::::: :::::
12257 GGGCCAGCCATGACGTGCGGACACGCGCTCGCAGGTGTTCATATCGG 12306
385 alTTPSerLysAsnValAsnGlnLeuValSerThrHisGlyTyrSerGln 401
    ::::: ::::: ::::: ::::: ::::: :::::
12307 CCGGTCCAAAGACCTCTCGAGCTGTCTCCACGACGCTACTCTGCGAG 12356
402 AsnGlnIleIleValTyrPArgTyrProThrMetSerLysLeuAlaThrLe 418
    ::::: ::::: ::::: ::::: ::::: :::::
12357 AACGAGATATCTGTGGAATATCCCTCCGACGCAAGTGGCCAACT 12406
418 uThrGlyHisThrTyrArgValLeuTyrLeuAlaIleSerProAspGlyG 435
    ::::: ::::: ::::: ::::: ::::: :::::
12407 GACGGCCATTCGTATCGTGTCTCTATCTGCGCTGATGTCGATGTTG 12456
435 1nThrIleValThrGlyAlaGlyAspGlyuThrLeuArgPheThrPasnVal 451
    ::::: ::::: ::::: ::::: ::::: :::::
12457 AGCTATTGTTRAGGCGCGCGCGACGAGACGCTGCGATTGTGGACACTA 12506

```



```

2882 CAGGTGGCCATTGGACACCCAGTCCGGGTACGTGACCATTTGGATGCCGA 2931
2931 aargcylslyslsllleargsermetgluglhisargleuargvalgla 256
2932 GAACCAAAAGCAGATAAACGAGTGAAGAGCAGCTGGCGGCGGTACCG 2981
256 laleualatrpserserleuileuenserllyglarapslyasn 272
2982 CGCTACTTGGTGGCGCAACCGATGCGCAGTGGCTCCGCGACCCCTCA 3031
273 lletyglarapsllleargtrhnglualsphevalserlys...le 268
3032 ATTTCAGCGGGATATCCGCAATCCGCCAGCACATTAACCCGCGCTT 3081
288 userglyhslslysergluvalcysglyleuylstrpsertyraspa 305
3082 GCGCGGACCAAGACTAGAGGTGTGGCTTACAGTGTGCCCGACGAAC 3131
305 rgluleualaserglyglasnaspsnlyslleuphevaltrpsn 321
3132 GATACTTGGCCAGCGGTGGCAGCAGACACCGCTGTGGTGTGACCCG 3181
322 Hlssetrhglinprovalleuylstrcysgluhisthralalaval 338
3182 GATTGGCCGAGCCCATTTACCCCTTCGACGACACAGACGCGGTCA 3231
338 salalialatrpserprohisleuhsiglyleuualaserglylg 355
3232 GCGCGTGGCGGTGGTGGCCACATAGCCGCTTCTGGCCAGTGGCGG 3281
355 lythralasparcysileargpmetrpasrthtrthrasnserhis 371
3282 GGTCCGCGATGCTGCTGCTGCTTGGAAAGCTCCTACCGGAAAGCTG 3331
372 leuSerCysMetaspThrglysergluvalcysasnleuvaltrpser 388
3332 GTGAAGTGCATCAACACGCGGCCACAGATACCAACACGCGTGGCCAG 3381
388 sasnlvalasngluleuvalserthhisgltyrsergluasn 405
3382 GGATTCGCGAGAGCTGTACACCCAGCAGCAGCAGCCAGCCAGCTCA 3431
405 levaltrpartrpProthmetserlysleualatrhleuThrglylis 421
3432 TTGCTGGCGGCTATCCTCGCTGAAGCAGATGGCCAGGCTGTCCGGGAC 3481
422 ThrYrArGValLeuYrLeuAlaIleSerProaspGlyInThrIleVa 438
3482 ACACAGCGGTACTCCACCTGTCACTGATCGGATACGATCATTACT 3531
438 lThrglylaGlyaspGluThrleuArgPheTrpAsnValPheProSer 454
3532 GACTGGCGCGCAGATGAGACGCTTGGCTGACCGCTTTTACCAAAC 3581
455 .....ProlySerGlnAsn..... 459
3582 AGAAATATTCAAAGAGCCAAACCTCGCTCCGTTTGAAGCATTA 3631
460 ThGluSerGluIleGlyAlaLeuSerLeu 469
3632 AGATGAGAGCAAAATGCGACGTATCTTTG 3662
seq_name: /STDSL/gcgdata/mol-d-geneseq/geneseq-emb1/NA2001B.DAT:ABL18423
seq_documentation_block:
ID ABL18423 standard; DNA; 1369 BP.
XX
AC ABL18423;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6742.
XX

```

```

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
OS Drosophila melanogaster.
PN WO200171042-A2.
PD 27-SEP-2001.
PE 23-MAR-2001; 2001WO-US09231.
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PA (PEKE ) PE CORP NY.
PI Venter JC, Adams M, Li PWD, Myers EW;
DR WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 6742; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL1840-ABL16175) and the encoded proteins
XX (AB1857737-AB1872072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1369 BP; 304 A; 394 C; 394 G; 277 T; 0 other:
XX
XX alignment_scores:
XX      Quality: 995.50      Length: 441
XX      Ratio: 3.211      Gaps: 9
XX      Percent Similarity: 70.295      Percent Identity: 45.805
XX
XX alignment_block:
XX US-09-701-572-2 x ABL18423 ..
XX
XX Align seq 1/1 to: ABL18423 from: 1 to: 1369
XX
XX 52 ArgPheIleProSerArGserAlaSerlysPheAlaLeuPheAspIleAs 68
XX ||| |||||:||||:||||:||||:||||:||||:||||:||||:||||: 68
XX 83 CGCTGCATTCGCCACAGCAGCTGTAAACCACTGGAACCGCACTTTGTCA 132
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 132
XX 68 nThrProthnglulYrArgAspAspSerSerSerAlaYrThrTrhLeu 85
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 85
XX 133 TGTCTCCGGAGACTCCATGTGAGACAGAGCTGTGTGATCATGCTCT 182
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 182
XX 85 euArgthrAlaLeuPheGlyProAspValAlaGlyProValThrProGlu 101
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 101
XX 183 TGGGCAC..... 193
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 193
XX 102 LysThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArgTy 118
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 118
XX 194 TTCCTGAGCAGAGACATCACTTCCTTGGAGAAC..... 226
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 226
XX 118 rlystrhnglulnhrArgInSerMetHisSerLeuSerProPheMetAspa 135
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 135
XX 227 ....TSCGAGCAGACCGCAGAGTCCATGCC.....CAGACTG 260
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 260
XX 135 spAspPheValProGlyValAsnHis..... 143
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||: 143

```

```

261 ATGATTTTAGGATGCGCATACCGTATCTTGTGTTGCCAGCGTTCT 310
144 ... SerProValLysAlaPro..... 149
311 TCGTGCCGAGTACGACACCATTTCCCAACATCCAGAGAGCCAGCG 360
150 ..... ArgLysValProArgSerProT 157
361 GGTGCTCAGCATCTTCAGAGACCCATGCGCCGACTGCCAGCTGCCCT 410
157 yLysValLeuAspAlaProAlaLeuGlnAspPheTyrLeuAsnLeu 173
411 ACAAGTACTCATGCGCGGAGCTGCAAGAGACTTCTACTGAACTT 460
174 ValAspTyrSerSerHisAsnValLeuAlaValGlyLeuGlyAsnGly 190
461 ATGAGTCTGCTCGTCAAAACACCTGCGCTGCTGCTGCTGCTGCTGCT 510
190 LtyrLeuTyrPasnAlaCysSerSerLysValThrLysLeuCysAspLeu 206
511 TTTCTCTGAGTCCGCTCAGTGTGCTAGCTGACGCGACTATCCGACTTCA 560
207 ..GlyValAspAspCysValCysSerValGlyTyrPalaGlnArgLysThr 222
561 ACAATGAGATTAACCTGATTACAGCAGTTCAGCTGCAGCGTGAAGCCG 610
223 HisLeuAlaValGlyThrAsnAsnGlyLysValGlnIleTyrPasnAla 239
611 CAGGTGCGCATTTGGCACCAGTCCGGTACGTACATTTGGGATGCCGA 660
239 aArgCysLysLysIleArgSerMetGlnGlnIleArgLeuArgValGly 256
661 GAACCAAAAGCAGATTAACCGACTGAGACAGCATCGCGCGTGTACCG 710
256 LalaLeuAlaTyrSerSerSerLeuLeuSerSerGlyLysArgAspLys 272
711 CGTACTCTGTGTGCGGCAACCGATTGCGCAGTGTGCGCAGCGCTCA 760
273 IleTyrGlnArgAspIleArgThrGlnGlnIleAspPheValSerLys... 288
761 ATTTTTCGCGGAGATTCGCAATCCCGCAGCCAGCATATACCGCGCTG 810
288 uSerGlyHisLysSerGlnValCysGlyLysLysTyrAspAsn 305
811 GCGCGGACACAGCTAGAGGTGTGCGCTACAGTGTGCGCAGCAGCAATC 860
305 rgluLeuAlaSerGlyGlyAsnAspAsnLysLeuPheValTyrPasnGln 321
861 GAATCTTGGCCAGCGGTGCGCAGCAGCAGCTGTGTTGTTGACCGAC 910
322 HisSerThrGlnProValLeuLysTyrCysGlnHisThrAlaAlaVal 338
911 GATTGGCGGAGCCCATTTACGCTTCGACGACACAGCGGTGTGCA 960
338 sAlaIleAlaTyrSerProHisLeuHisGlyLeuAlaLeuAlaSerGly 355
961 GCGCTGTGCTGCGCAGCATTAAGTCCGCTTCTGCGCAGAGTGGCGT 1010
355 LyrThrAlaAspArgCysIleArgPheTyrPasnThrThrTranSerHis 371
1011 GGTGCGCGCATCGCTGCTGCTGCTGCAACGCTTCACCGGAGAGCTG 1060
372 LeuSerCysMetAspThrGlySerGlnValCysAsnLeuValTyrSer 388
1061 GTAGAGGATCAACACACCGGCGCCAGATCAGCACTGCGTGGCGCAG 1110.
388 sAsnValAsnGlnLeuValSerThrHisGlyTyrSerGlnAsnGlnIle 405
1111 GGAATCCCGAGAGCTGTGTTACACACGACGACGCGCAGCAGGTCA 1160
405 LeValTyrArgTyrProThrMetSerLysLeuAlaThrLeuThrGlyHis 421
1161 TTGCTGTGCGCTATCCCTCGCTGAAGCAGATGGCAGGCTGTCCGGGCG 1210

```

```

422 ThrTyrArgValLeuTyrLeuAlaIleSerProAspGlyGlnThrIleVal 438
1211 ACACAGCGGGGACTCCACGTGTCAGTGCAGTCCGATTAACGATCCATAGT 1260
438 LThrGlyAlaGlyAspGlnThrLeuAlaGlyPheTyrPasnValPheProSer 454
1261 GACTGCGCGCGCAGATGAGAGCGTTCGCTTGTGACCGTTTACCAAC 1310
455 ..... ProlSerGln 458
1311 AGAAATATTCAAAAGACCA 1331
seq_name: /sids1/gcgdata/hold-geneseq/gene-seq-emb1/NA2001B.DAT:ABL07411

```

seq_documentation_block:

ID ABL07411 standard; cDNA; 1706 BP.

XX ABL07411;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide seq ID NO 16715.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

XX P-PSDB: ABB63308.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 16715; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

XX sequences (AB101840-AB116175) and the encoded proteins

XX (AB157737-AB172072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 1706 BP; 432 A; 446 C; 458 G; 370 T; 0 other;

alignment_scores:

Quality: 917.00 Length: 478

Ratio: 2.866 Gaps: 10

Percent Similarity: 66.946 Percent Identity: 40.377

alignment_block:

US-09-701-572-2 x ABL07411 ..

Align seg 1/1 to: ABL07411 from: 1 to: 1706

```

42 ThrSerProSerArgThrIle...TyrSerAspArgPheIleProSerAr 57
1176 AGCGAGGCAACGACAAATCTGTGTATGTCGGCGGACCGGTGG 1225
354 ACCAGCCCTCTTAAGACGCCAGGAGCGGATCGCTTATTCGGAATG 403
57 gSerAlaSerLysPheAlaLeu.....PheAspIleAsnThrProThg 72
404 GCGGCTACCAACTTGTAGTACACACTTGTGTGTAACAAGAAGACTCG 453
72 LuGlyArgAspAspSerSerSerAlaTyrThr..... 83
454 GCGATAGTCGATGAGGAGAACAGACAGCCACTCGACAAACAGCAAC 503
84 .....Le 84
504 GAGAGCAATGTCAGGCTCGCTCCACAGGCGGACCGGACAAACTAT 553
84 uLeuArgThrAlaLeuPheGlyProAspValAlaGly..... 96
554 CTCTGAAGTGGCCAGCTCGGTACTCCAAAGGCGGCGCATTTGTGCT 603
97 .....ProValThrProGluLysThrAspSerProSerMet 108
604 ACCAAACAGAGCTCCGCTGCTCCAGAAACACACACAACTCCCTGAG 653
109 ThrLeuProAsnArgAsnIlePheArgTyrLysThrIleArgGlnSe 125
654 GTCGTG.....TACTCATTTAAGACCCCATATTC 682
125 rMetHisSerLeuSerProPheMetAspAspPheValProGlyVala 142
683 CACAAAGACTGGCTCA..... 698
142 snHisSerProValLysAlaProArgLysValProArgSerProTyrLys 158
699 .....CGCTATATACCCACACACACACACCGGAGAG 725
159 ValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuVala 175
726 ATTCTGATGACACCTGATTTTATTAACGATTACTTAAATCTTATGGA 775
175 pTyrSerSerHisAsnValLeuAlaValGlyLeuGlyAsnGlyValTyrL 192
776 TTGGAGTCCGCAATATATGTGGCTGTGGCTGTGGAGTTCGCTATG 825
192 euTrpAsnAlaCysSerSerLysValThrLysLeuGlyAspLeuGlyVal 208
826 TGTGAAAGCAGACACCGGAAATATCGACAGCTTACGGAGTTTGAGAG 875
209 AspAspCysValCysSerValGlyTrpAlaGlnArgGlyThrHisLeuAl 225
876 GCGGACTACGAGCTCGTATCGTATCGAGATCCAGAGGCGGACGACTTC 925
225 aValGlyThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCysL 242
926 CATCGGCAACGACACCGGCTGCGGAGCTGTGGAGTCTGCTCCAAAGTA 975
242 yLysIleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAla 258
976 AGGCTCTCGAGTATGATGATGACACAGTCCGAGTGCGATTCCTTGCC 1025
259 TrpSerSerLeuLeuSerSerGlyLysArgAspLysAsnIleTyrGly 275
1026 TGAAGTCAATCTGTTCTCTGCGACCGCGGATGGACACATTTGCCA 1075
275 nArgAspIleArgThrGlnGluAspPheValSerLysLeuSerGlyHisL 292
1076 CCAGGATGTGTCAGCTGAGCAAACTTTCACATTTCCGAGACACA 1125
292 ySerGluValCysGlyLeuLysTrpSerTyrAspAsnArgGluLeuAla 308

```

```

1126 GCGAGAGGTTTGGCGCCTAAAGTGTCCAGGATTCATATTGCT 1175
309 SerGlyGlyAsnAspAsnLysLeuPheValTrpAsnGlnHisSer.... 323
1176 AGCGAGGCAACGACAAATCTGTGTATGTCGGCGGACCGGTGG 1225
324 .....ThrGlnProValLeuLysTyrCysGluHisThrAla 336
1226 CGTGGAGACTGCCACGATCTCTTGCACAAATTAACAGCATTCAAGCTG 1275
336 lavalysAlaIleAlaTrpSerProHisLeuHisGlyLeuLeuAlaSer 352
1276 CAGTGTGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1325
353 GlyGlyGlyThrAlaAspArgCysIleArgPheTrpAsnThrTrpAs 369
1326 GAGGCGGCGACCGCGGATCGCTGCATCAAGTTCGAAATGGAACATGG 1375
369 nSerHisLeuSerCysMetAspThrGlySerGlnValCysAsnLeuVal 386
1376 CACTTATATGAATATCCGAGACTCCAAAGTCCGAGCTGTCTCTGCTCT 1425
386 rPserLysAsnValAsnGluLeuValSerThrHisGlyTyrSerGlnAsn 402
1426 TTCTCGGCACTACAGAGAGCTGATCTGTGGAGTGTTCGTAACAC 1475
403 GlnIleIleValIleTrpArgTyrProThrMetSerLysLeuAlaThrLeu 419
1476 CAACTGACCAATTTGGAATATCCCAACAAATGTGAGAACACCACTTTGAC 1525
419 rGlnHisThrTyrArgValLeuTyrLeuAlaIleSerProAspGlyLys 436
1526 TGGACACACGACGACGATCTCCAGATGCCATGTCTCCGAGCGGACCA 1575
436 hrIleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnValPhe 452
1576 CAGTATCAGCGCGGAGCTGATGAATACCTGCTGTTGAGACTGCTTC 1625
453 ...ProSerPro.....LysSerGlnAsnThrCysLeuGlu 464
1626 GCTCCGATCCGTTGGCGCTGCAAGAGGCAAGTTCGACGACGACGAG... 1670
464 eGlyAlaLeuSerLeuGlyArgThrThrIleArg 475
1671 .GGCAACAGAGCGGTGTCGACAGACATCCGT 1703
seq_name: /std1/gcgdata/hold-geneseq/geneseg-nbml/NA195.DAT:AAQ96099
seq_documentation_block:
ID AAQ96099 standard; cDNA; 1767 BP.
XX
AC AAQ96099;
XX
DT 23-JAN-1996 (first entry)
XX
DE
XX Sequence encoding rat cell cycle protein p55CDC.
XX
XX Cell cycle; protein; p55CDC; antibody; chemotherapy; modulation;
XX inhibition; growth; ss.
XX
XX Rattus rattus.
XX
OS
XX
FH Key location/Qualifiers
FT CDS 193..1692
FT /tag= a
FT /product= Cell cycle protein p55CDC.
XX
PD 17-AUG-1995.
XX
XX 13-FEB-1995; 95WO-US01806.
XX

```



```

61 sphenalaurheaspilleasnthrprothrcluclyargaspaspsers 78
   ::::: ::::: ::::: ::::: ::::: :::::
377 GATGGAGGTGGCAGCTTCTCTGAGCAGAGAACCACTGTAACCA 426
78 erSerAlaTyrThrThrLeuValArgThrAlaLeuPheGlyProAspVal 94
   ::::: ::::: ::::: ::::: ::::: :::::
427 GCCAG..... 431
95 AlaGlyProValThrProGlyLysThrAspSerProSerMetThrLeuR 111
   ::::: ::::: ::::: ::::: ::::: :::::
432 .....ACGCCCAACAGAGAGACATCAGAAAGCCTGGGCTTT 469
111 oasnargasn.....IlePheArgTyrLysT 120
   ::::: ::::: ::::: ::::: ::::: :::::
470 GAACCTGAACGGTTTGTATGAGAGAACCAAGATCTTGCTGCTCAGT 519
120 hrGlyThrArgGlnSer.....MetHisSerLeuSerProPhe 132
   ::::: ::::: ::::: ::::: ::::: :::::
520 GAAACCAACCAAAATGCGCAGAGGTATTCAGAACAGACTGAAGTACTC 569
133 MetAspAspAspPheValProGlyValAsnHisSerProValLysAlaPr 149
   ::::: ::::: ::::: ::::: ::::: :::::
570 TACACCCAAAGGCCACTCTGGC.....TCCACCCGAGAGACCTG 610
149 ArgGlyValProArgSerProTyrLysValLeuAspAlaProAlaLeuG 166
   ::::: ::::: ::::: ::::: ::::: :::::
611 CCGTTACATTCCTTCCCTGCGCAGACCGTATCTGATCGCTGAAATCC 660
166 InAspAspPheTyrLeuAsnLeuValAspTrpSerSerHisAsnValLeu 182
   ::::: ::::: ::::: ::::: ::::: :::::
661 GAAATGACTATTAACCTGAACTTGAGTGTGAGTTCGGAATGTACTG 710
183 AlaValGlyLeuGlyAsnCysValTyrLeuThrAsnAlaCysSerLeu 199
   ::::: ::::: ::::: ::::: ::::: :::::
711 GCGGTGGACTGGACACAGCTGTACTGTGAGTGCACACTCTGGTGA 760
199 sValThrLysLeuCysAspLeu...GlyValAspAspCysValCysSerV 215
   ::::: ::::: ::::: ::::: ::::: :::::
761 CATCTGGAGCTTTGCAATGAGCAGCCTGGGCAATATATATCCCTGT 810
215 alGlyThrAlaGlnArgGlyThrHisLeuAlaValGlyThrAsnAsnGly 231
   ::::: ::::: ::::: ::::: ::::: :::::
811 TGGCCTGGATCAAAAGAGGCAACTTGTGGTGGCAGCAGCAGCTCT 860
232 LysValGlnIleTrpAspAlaAlaArgCysLysLysIleArgSerMetG 248
   ::::: ::::: ::::: ::::: ::::: :::::
861 GAGGTGCAGCTATGGATGTGCAGCAGCAAAACGCTTCCAATATATAC 910
248 uGlyHisArgLeuArgValGlyAlaLeuAlaTrpSerSerSerLeuLeu 265
   ::::: ::::: ::::: ::::: ::::: :::::
911 CAGTCATCTGGCCGAGTGGCTCCCTAAGCTGAACAGCTATATCCCTGT 960
265 erSerGlyGlyArgAspLysAsnIleTyrGlnArgAspIleArgThrGln 281
   ::::: ::::: ::::: ::::: ::::: :::::
961 CCAGTGGTTCAGTCTGGCCACATCCACACCATGATGTCGGGTACCA 1010
282 GluAspPheValSerLysLeuSerGlyHisLysSerGlyValLysGlyLe 298
   ::::: ::::: ::::: ::::: ::::: :::::
1011 GAACACCATGTGGCCACACTGAGTGGCCACACCGAAGAGTGTGGGCT 1060
298 uLysTrpSerTyrAspAsnArgGluLeuAlaSerGlyLysAsnAspAsn 315
   ::::: ::::: ::::: ::::: ::::: :::::
1061 GCGGTGGGCCCCAGATGAGCAGCATTTGGCAGTGTGATATGATACT 1110
315 yLysLeuPheValTrp.....AsnGlnHisSerThrGlnProVal 327
   ::::: ::::: ::::: ::::: ::::: :::::
1111 TGGTCAATGTGTGGCTAGTGTCTCTGAGAGAGGTGGTGGCTCTCTG 1160
328 LeuLysTyrCysGlnHisThrAlaAlaValLysAlaIleLeuTrpSerPr 344
   ::::: ::::: ::::: ::::: ::::: :::::
1161 CAGACATTCACCCACATCAAGAGGCTGTCAAGGCTGATAGATGTGTCC 1210
344 HisLeuHisGlyLeuAlaSerGlyGlyThrAlaAspArgCysI 361

```

```

1211 CTGGCAGTCCCATGTCTCTGGCAACAGAGGGGCGACCATGATCCGACAA 1260
361 leArgPheTrpAsnThrThrThrAsnSerHisLeuSerCysMetAspThr 377
1261 TTCGATCTTGAAATGTGTGCTGTGGGGCGCTGTGTGAGTCCGCTGATGCC 1310
378 GlySerGlnValCysAsnLeuValTrpSerLysAsnValAlaGlnLeuVal 394
1311 CATTCGAGGTGTGTCTCATCTGTGTCTCCCATTCACAGAGAGTCTCAT 1360
394 lSerThrHisGlyTyrSerGlnAsnGlnIleLeuValTrpArgTyrProt 411
1361 CTCAGGCGCATGGCTTGCACAGAACCAAGCATATTGGAATACCA 1410
411 hrMetSerLysLeuAlaThrLeuThrGlyHisThrTyrArgValLeuTyr 427
1411 CCATGGCCAGGCTGCTGAACTCAAGGTCACACATCCGCGCTCAGT 1460
428 LeuAlaIleSerProAspGlyInThrIleValThrGlyAlaGlyAspG 444
1461 CTGACCATGAGCCAGATGGGCGCCACAGTGGCATCCGACAGCAGATGA 1510
444 uThrLeuArgPheTrpAsnValPheProSerProLysSerGlnAsnThrG 461
1511 GACCTGAGGCTATGGGCGCTTTGTAGTGTGACCTGGCGGCGGCGG 1560
461 luserGlu 463
1561 AGCGGGAG 1568
seq_name: /sids1/gcdata/hold-geneseq/geneseg-emb1/NA2001B.DAT.ABL07410

```

seq_documentation_block:

```

ID ABL07410 standard; CDNA; 3852 BP.
XX
AC ABL07410;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 16712.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacetical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR P-PSDB; ABB63307.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 16712; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

```


|||||
2844 ATCCGT 2849

seq_name: /SIS1/gcgdata/hold/geneseq/geneseqn-emb1/NA2000.DAT:AAc46256

seq_documentation_block:

ID AAC46256 standard; DNA; 2697 BP.

XX AAC46256;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 49475.

DE Hybridisation assay: genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

OS EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0128714.

XX 19-APR-1999; 99US-0129845.

XX 21-APR-1999; 99US-0130077.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 18-MAY-1999; 99US-0134221.

XX 19-MAY-1999; 99US-0134370.

XX 20-MAY-1999; 99US-0134941.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 17-JUN-1999; 99US-0139453.

XX 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139464.

PR 21-JUN-1999; 99US-0139763.

PR 22-JUN-1999; 99US-0139817.

PR 23-JUN-1999; 99US-0140353.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 20-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145218.

PR 26-JUL-1999; 99US-0145224.

PR 27-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.


```

357 aaSPaRGcysIleArgPheTrpAsnThrThrAsnSerHisLeuSerC 374
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1599 AGACAGAGAGATTAAGTTCTGGAATCTACACACTGGCGCTTCTGAAT 1648
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 ysmetsphrIglYserGlnValCysAsnLeuValTrpSerLysAsnVal 390
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1649 CAGTAGACACTGGTCCCAAGTTGCTGTGTATGAGCAAGAATGAA 1698
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
391 AsnGlnLeuValSerThrHisGlyTrpSerGlnAsnGlnIleLeuValTr 407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1699 AGAGATTGCTTACGTCACATGGGTTTACACAGAACTACCTTACCTTGG 1748
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
407 pArgTrpProThrMetSerLysLeuAlaThrLeuThrGlnHisThrTYA 424
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1749 GAAGTACCATCATGCTGAAATGGCTGAGCTCACTGGCTCATCATCA 1798
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
424 rGValLeuTrpLeuAlaIleSerProAspGlyInThrIleValThrGly 440
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1799 GAGTTCTATATATAGCCAGAGTCCAGATGTTGACTGACTTACCTCAGCA 1848
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 ALAGLYAspGluThr.....LeuArgPheTrpAs 450
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1849 GCAGGAGAGAGACCCCTGAGCCCAACGCAAGCTTTGCTTTGGCT 1898
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
450 nValPhePro 453
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1899 TTTAAACCT 1908

```

seq_name: /SISL/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.ABL10086

seq_documentation_block:

ID ABL10086 standard; cDNA; 14482 BP.

AC ABL10086;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 24740.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WC200171042-A2.

XX 27-SEP-2001.

PE 23-MAR-2001; 2001MO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB65983.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Claim 1; SEQ ID NO 24740; 21bp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB057737-AB072072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SO Sequence 14482 BP; 4240 A; 2898 C; 2950 G; 4394 T; 0 other;

alignment_scores:

Quality: 619.50 Length: 237
Ratio: 3.367 Gaps: 4
Percent Similarity: 77.637 Percent Identity: 51.055

alignment_block:

US-09-701-572-2 x ABL10086 ..

Align seg 1/1 to: ABL10086 from: 1 to: 14482

```

250 HisArgLeuArgValGlyAlaLeuAlaTrpSerSerLeuLeuSer 266
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 CACAGTCCCGATGGATCTTGGCTGGAACATCTCTGTTCTC 51
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
266 rGlyGlyArgAspLysAsnIleTyrglnArgAspIleArgThrGlnGlu 283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 TGCAGCCGGGATGGCACATTGTCACACAGATGTCGTGACGTGAGC 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
283 sprheValSerLysLeuSerGlyHisLysSerGluValCysGlyLeuLys 299
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 ACAAGCTTCCACATGTCGGACACACAGAGAGATGTCGGCCATAAA 151
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 TrpSerTyraSpAsnArgGluLeuAlaSerGlyGlyAsnAspAsnLys 316
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 TGGTCCACGAGATTTCAGATATTTGGCTTACGGAGGACACACATCTGG 201
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
316 upheValTrpAsnGlnHisSer.....ThrGlnPro 327
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 GAATGTTGGTGGCGGCGGCGGCGGCGGCGGAGACTGCACCGATCCCT 251
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
327 alleuLysTyrcysGlnHisThrAlaValLysAlaIleAlaTrpSer 343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 TGCACAAATTCACAGACATCAAGCTGCACTGAGTGGCTTGGCTGTGT 301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
344 ProHisLeuHisGlyLeuAlaSerGlyGlyGlyThrAlaAspArgCys 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 CCTGGGCAACCAAGTACTAGCTCTGAGCGGCGGACCGCGATGCGTG 351
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
360 sIleArgPheTrpAsnThrThrAsnSerHisLeuSerCysMetAspTr 377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 CATCAAGTTCTGGAATGTGAACATGGCACTTTAATGAATCCGTGACT 401
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
377 hrcLysSerGlnValCysAsnLeuValTrpSerLysAsnValAsnGln 393
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 CCAAGTCGACAGTCTGCTCTCTCTCTTTCCTGCGACACACAGAGACTG 451
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
394 ValSerThrHisGlyTrpSerGlnAsnGlnIleLeuValTrpArgTrp 410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 ATCTTGCGCATGTTTGTCTACACCAACATGACCATTTGGAATATACC 501
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 oThrMetSerLysLeuAlaThrLeuThrGlnHisThrTYAArgValLeu 427
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 AACAAATGTGAACAGCCGATTGACTGACACACAGCTCACAGATTCTCC 551
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
427 yreValIleSerProAspGlyGlnThrIleValThrGlyAlaGlyAsp 443
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
552 AGATGGCATGTCTCCGACGCGACAGCATGATCAGCCCGGAGCTGAT 601
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
444 GluThrLeuArgPheTrpAsnValPhe...ProSerPro..... 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
602 GAACCTCTGCTGTTGAAAGTGTCTGCTCCGATCCGTTGGCGGTCA 651
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
456 LysSerGlnAsnThrGlnSerGluIleGlyAlaLeuSerLeuGlyArgT 472
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

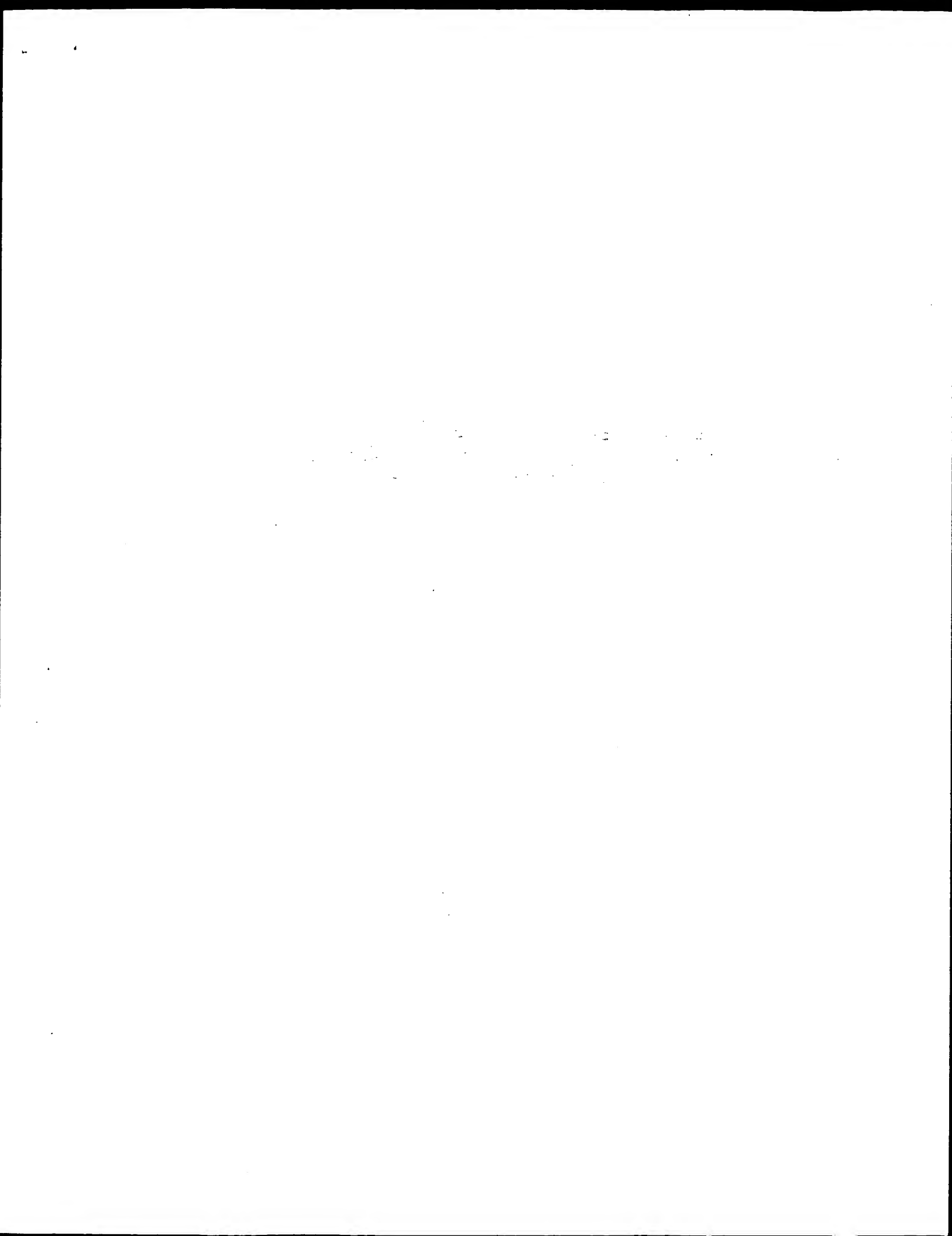
```

Wed May 29 14:59:19 2002

us-09-701-572-2.rng

Page 21

652 GAAGCAGTTTCAGCCAGCAG.....GCCAACAGAGCGTGTTCGAC 695
472 hThtlLeatg 475
:::|||||
696 AGAGCATCGT 706




```

17 pasSerProProgluProSerProgluSerLeuArgHisValSerA 34
123 TAAATTCACACCGCCGAGACCATCACCAGAGCTCCGCCCATGTAGCC 172
34 rgmellleasnSerasnHisTyrThrSerProSerArgThrIleTyrSer 50
173 GTATGATCAACAGCACCATTTACACTCACCCTTCTGCACAACTACTACC 222
51 AsparagheilleProSerArgSerAlaSerLysPheAlaLeuPheasp11 67
223 GATAGGTGATCTCCAGTACATCTGCTTCCAAATGCGCTTGTGATAT 272
67 easnThrProThrGluGlyArgAspAspSerSerSerAlaTyrThrThr 84
273 CAATACCTCCGACGAGAGACGCGCATGATATTCACAGCGCTTATACACATC 322
84 euleuArgThrAlaLeuPheGlyProAspValAlaGlyProValIthrPro 100
323 TTCTTACACAGCGGCGTGTGTTGACCGCATGCTGCAGTCCGCGTTACGCGG 372
101 GluThrAspSerProSerMetThrLeuProAsnArgAsnIlePheArg 117
373 GAGAAACCGACCTCGCGGTGATGACATTTGCCAATAGGAATATTTTAg 422
117 gTyrLysThrGluThrArgGlnSerMetHisSerLeuSerProPheMetA 134
423 GTATTAAGACGAGACGAGACGACATGCATGCGCTTTCGCGCTTATAG 472
134 spAspAspPheValProGlyValAlaSerHisSerProValLysAlaProArg 150
473 ATATGATTTTGTCTCGTATTAATCATAGTCCGCTTAAAGGCTCTTAG 522
151 LysValProArgSerProTyrLysValLysAspAlaProAlaLeuGlnAs 167
523 AAGTTCCTCGATCGCCTTATAGCTTTTGATGACACCTTCCTTGCACAA 572
167 pasPheTyrLeuAsnLeuValAspTrpSerSerHisAsnValLysAlaVal 184
573 TGATTTTATCTGAATCTGTAGATGTTGCTTACACAAATGTTGGCTG 622
184 alGlyLeuGlyAsnGlyValTyrLeuTrpAsnAlaCysSerSerLysVal 200
623 TTGGTTGGGTAACTGTGTTATCTCTGGAATGCTGTGAGCAGCAGGA 672
201 ThrLysLeuCysAspLeu.GlyValAspAspCys.ValCysSer.ValG1 216
673 ACTAATATTATGTCATTTGGGGGGGTGATGCAATGGGTTTCTCTGTTGC 722
216 yTTPAlaGlnArgGlyThrHisLeuAlaValGly 227
723 TTGGGCTCAACGCGGTACTCATCTGCTGTGGA 756
seq_name: gb_est2:BI270288
seq documentation block:
LOCUS BI270288 630 bp mRNA linear EST 18-JUL-2001
DEFINITION NE006A08FL1065 Developing flower Medicago truncatula cDNA clone
ACCESSION BI270288
VERSION BI270288.1 GI:14877742
KEYWORDS EST.
SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE
1 (bases 1 to 630)
Torres-Jerez, I., Scott, A. D., Harris, A. R., Gonzales, R. A., Bell, C. J.,
Flores, H. R., Imman, J. T., Weller, J. W. and May, G. D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula flower library

```

```

JOURNAL
COMMENT
Unpublished (2001)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 630 Std Error: 0.00
Plate: 006 row: A column: 08
Seq primer: TCACACGAGAAACAGCATATC.
FEATURES
source
1..630
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF006A08FL"
/clone_1ib="Developing flower"
/tissue_type="Developing flowers"
/dev_stage="Developmentally pooled. Contains a mixture of
very young, developing, fully-opened flowers and flowers
in early transition into pods."
/note="Vector: Lambda Zap; cDNA was prepared from polyA+
enriched, pooled samples of equivalent amounts of total
RNA from very young, developing, fully-opened flowers and
flowers transitioning into pods. The cDNA was
directionally ligated into the Uni-Zap XR vector
(Stratagene) and packaged using the Gigapack III Gold
packaging extracts. Phagmids containing cDNA inserts were
in vivo excised from the recombinant Uni-Zap XR vector
using Exassist helper phage and the E. coli strain
XL1-Blue MRF' (Stratagene). Excised plasmids were plated
using SORR cells."
BASE COUNT 159 a 97 c 167 g 204 t 3 others
ORIGIN
alignment_scores:
Quality: 1122.00 Length: 209
Ratio: 5.420 Gaps: 0
Percent Similarity: 99.043 Percent Identity: 98.565
alignment_block:
US-09-701-572-2 x BI270288
Align seg 1/1 to: BI270288 from: 1 to: 630
112 AsnArgAsnIlePheArgTyrLysThrGluThrArgGlnSerMetHisSe 128
|||||
3 AATAGGAATATTTTATGATATGAATGATTTTGTCCGGAATTAATCATATGTC 52
128 rLeuSerProPheMetAspAspAspPheValProGlyValAlaSerHis 145
|||||
53 GCTTTCGCGCTTATGATATGATGATTTTGTCCGGAATTAATCATATGTC 102
145 roValLysAlaProArgLysValProArgSerProTyrLysValLysAsp 161
|||||
103 CGGTTAAGGCTCTCAGAGAGGTTCTCGATCGCCTTATAGGTTTGAT 152
162 AlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuValAspTrpSer 178
|||||
153 GCACCTGCTTGGCAAGATGATTTTATCTGTAATCTGTAGATTCGCTTC 202
178 rHisAsnValLeuAlaValAlaGlyLeuGlyAsnGlyValTyrLeuTrpAsn 195
|||||
203 ACACAAATGTGTGCTTGTGTTGGTAACTGTGTTATCTCTGGAATG 252
253 CTGTGACACAGCAAGGTAACTAAATTAATGATTTGGGGGTGATGATGTG 302
195 lacysSerSerLysValThrLysLeuCysAspLeuGlyValAspAspCys 211
|||||
212 ValCysSerValGlyTTPAlaGlnArgGlyThrHisLeuAlaValGly 228
|||||
303 GTTGTCTCTGTGTGGGCTCAACGCGGTACTCATCTGCTGTGGAAC 352

```

228 TAAAGATGTAAGTTCAGATTTGGATGACAGCAAGATGCAAGAAATAA 402
 333 TAAAGATGTAAGTTCAGATTTGGATGACAGCAAGATGCAAGAAATAA 402
 245 TAAAGATGTAAGTTCAGATTTGGATGACAGCAAGATGCAAGAAATAA 402
 403 TAAAGATGTAAGTTCAGATTTGGATGACAGCAAGATGCAAGAAATAA 402
 262 SerLeuSerSerGlyGlyArgAspLysAsnIleTyrGlnArgAspIle 278
 453 TCTCTTTGCTCTGCTGAGCGGATAGATATTTATTCACAGATAT 502
 278 eargthincinuasphevalserlyleuSerGlyHisLysSerGluV 295
 503 ACGCACACAAAGATTTTGTAGTAACTGTCAGACACAAATCAANAG 552
 295 aLysGlyLeuLysTyrSerTyrAspAsnArgGluLeuAlaSerGly 311
 553 TTTGTGAGCTAGATGTCATATATATACCGAGTTGGATCTGAGAGA 602
 312 AsnAspAsnLysLeuPheValTyrAsn 320
 603 AATGACACAAATGTTGTTGGAT 629

seq_name: gb_est2:BG044933

seq_documentation_block:

LOCUS BG044933 648 bp mRNA linear EST 28-NOV-2001
 DEFINITION saas3d02.1Y1 GM-cl059 Glycine max cDNA clone GENOME SYSTEMS CLONE
 ID: GM-cl059-939 5' similar to TR-09XED5 Q9XED5 CELL CYCLE SWITCH
 PROTEIN ; mRNA sequence.

ACCESSION BG044933

VERSION BG044933.1 GI:12492159

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE 1 (bases 1 to 648)

AUTHORS Shoemaker R., Klein P., Vodkin L., Eipelting J., Coryell V., Khanna

A., Bolla B., Marra M., Hillier L., Kucaba T., Martin J., Beck C.,
 Wylie T., Underwood K., Steptoe M., Theising B., Allen M., Bowers
 Y., Person B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk
 R., Ralster E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann
 R., Waterston R. and Wilson R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R./Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information

call: (800)-533-4363 or contact via email: c@resgen.com

High quality sequence stop: 405.

FEATURES

source

1. 648

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: GM-cl059-939"

/clone.lib="GM-cl059"

/tissue_type="whole seedling, 2 week old, etiolated"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
 XhoI; The cDNA library was constructed from mRNA isolated
 from 2 week old etiolated whole seedlings of P1468916.
 Complementary DNA was synthesized from mRNA using a primer

consisting of a poly(dT) sequence with a XhoI restriction
 site. EcoRI adapters were ligated to the blunt-ended cDNA
 fragments followed by XhoI digestion. The cDNA fragments
 were directionally cloned into the EcoRI- XhoI restriction
 site of the pluescript vector. The ligated cDNA fragments
 were transformed into DH10B host cells (Gibco BRL). This
 library was constructed in the laboratory of Dr. Randy
 Shoemaker at Iowa state university."

BASE COUNT 173 a 132 c 161 g 181 t 1 others

alignment_scores:
 Quality: 1111.00 Length: 216
 Ratio: 5.241 Gaps: 0
 Percent Similarity: 98.148 Percent Identity: 94.444

alignment_block:
 US-09-701-572-2 x BG044933 ..

Align seg 1/1 to: BG044933 from: 1 to: 648

227 GATGThrAsnAsnGlyValGlnIleTyrPaspAlaAlaArgCysLysIle 243
 1 GGAATGACATGTAAGTTCAGATTTGGATGACAGCAAGATGCAAGAA 50
 243 sileargsermetgluGlyHisArgLeuArgValGlyAlaLeuAlaTyrPS 260
 51 GATAAGATCATGAGAGGATCATGCGGTGAGCGGCTTGCGCTTGA 100
 260 eargserleuSerSerGlyGlyArgAspLysAsnIleTyrGlnArg 276
 101 GTTCATCTCTTTGCTCTGCGGCGAGGATAGATATTATCAAA 150
 277 AspIleArgThcIncuasphevalserlyleuSerGlyHisLysSe 293
 151 GATATCGTGACACAGAGATTTGTACGTAATGTGACGACACAAATC 200
 293 rgluValCysGlyLeuLysTyrSerTyrAspAsnArgGluLeuAlaSerG 310
 201 AGAGGTTTGGAGCTAGATGCGCTTATGATACGTCGATGGCATCTG 250
 310 lylGlyAsnAspAsnLysLeuPheValTyrAsnGlnHisSerThrGlnPro 326
 251 GAGGAATGACAAACAGATTTGTTGGAAATCAACACTCACTGACGCT 300
 327 ValLeuLysTyrCysGlnHisThrAlaValLysAlaIleAlaTyrSe 343
 301 GTCTGTAAGTACTGTGATGATACAGCAGCTGTAAAGCTATTGATGCTG 350
 343 rproHisLeuHisGlyLeuLeuAlaSerGlyGlyThrAlaAspArgC 360
 351 TCTGATCTTTCACGAGACTCTTGATCTGAGGAGAGAACTGACAGCCAT 400
 360 yslleargpPheTyrAsnThrThrAsnSerHisLeuSerCysMetLasp 376
 401 GCATACGTTTCTGGAATACACCAACTCACTTAAGCTGATGAGAC 450
 377 ThrGlySerGlnValCysAsnLeuValTyrSerLysAsnValAsnGluLe 393
 451 ACGGGAAGTCAGGTTTGCATCTGTCTGTCCAAAATGTCTAAATGAACT 500
 393 uvalSerThrHisGlyTyrSerGlnAsnGlnIleLeuAlaTyrPargTyrP 410
 501 ACTAAGACACATGCGTATTCACAGACAGATTTGTTGGATGACATACC 550
 410 rthMetSerLysLeuAlaThrLeuThrGlyHisThrTyrArgValLeu 426
 551 CCTCATGTCAAAAGTTGGCACTTACGCGGATCACTACAGATCTCT 600
 427 TyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaGly 442
 601 TATCTTGCATGTCTCCGATGCGAGACTATTGTACCTGGAGCTGGG 648

seq_name: gb_est2:BG444274

seq_documentation_block:

LOCUS BG444274 987 bp mRNA linear EST 15-MAR-2001

DEFINITION GA_Ea0023N11f Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum cDNA clone GA_Ea0023N11f, mRNA sequence.

ACCESSION BG444274

VERSION BG444274.1 GI:13353926

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Malvales: Malvaceae: Gossypium. 1 (bases 1 to 987)

AUTHORS Wang, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.

TITLE An integrated analysis of the genetics, development, and evolution of the cotton fiber

JOURNAL Unpublished (2000)

COMMENT Contact: Wang RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rtwing@clemson.edu

FEATURES

Seq primer: TAATACGACTCCTACTATAGG
High quality sequence start: 40
High quality sequence stop: 800.

source

1..987
location/Qualifiers

/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0023N11f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 258 a 203 c 256 g 265 t 5 others

ORIGIN

alignment_scores:

Quality: 1102.00 Length: 219
Ratio: 5.126 Gaps: 1
Percent Similarity: 98.174 Percent Identity: 94.064

alignment_block:

US-09-701-572-2 x BG444274 ..

Align seg 1/1 to: BG444274 from: 1 to: 987

254 VALGIALALEUALATPserSerleuSeuSerGlyAlaArgas 270
||| |||||:::|||||:::|||||:::|||||:::|||||
92 GCGAGGCGCTTATCTTGAGCTCTTCTCTGTTATCTTCTGCTGATGCA 141
270 physAsnIleTyrglnArgAspIleArgThrGlnGlnuAspPheValSerL 287
|||||:::|||||:::|||||:::|||||:::|||||
142 TAAGACATCTTCAAGAGATATACGTCGACAGATGATTTTCTAGTA 191
287 ylsuSerGlyHisLysSerGluValCysGlyLeuLysTrpSerTyraSp 303
|||||:::|||||:::|||||:::|||||:::|||||
192 AACTCTCTGTCACAGAGTCAGAGCTTGTGGCTGAGTGGCTTATGAC 241
304 AsnArgGlnLeuAlaSerGlyValAsnAspAluLysLeuPheValTrpAs 320
|||||:::|||||:::|||||:::|||||:::|||||
242 AACCGGAGCTGACATCTGCGAATGACAACTTTTGTGGACA 291

320 nGlnHisSerThrGlnProValIleuLysTyrcysGlnHisThr AlaAla 336
|||||:::|||||:::|||||:::|||||:::|||||
292 TCACATTCACACCCACCTGTTCTTAATACTGTGACATFACGCTGCC 341
337 ValIysAlaIleAlaATPserSerProHisLeuHisGlyLeuLeuAlaSerG 353
|||||:::|||||:::|||||:::|||||:::|||||
342 GTAAAGCAGATGCTTGCTGCTCCGATCTGCATCTGACGCTTCTTCATCTGG 391
353 yGlyGlyThrAlaAspArgCysIleArgPheTrpAsnThrThrThrAsnS 370
|||||:::|||||:::|||||:::|||||:::|||||
392 GGGGTGTACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 441
370 eHisLeuSerCysMetAspThrGlySerGlnValCysAsnLeuValTrp 386
:::|||||:::|||||:::|||||:::|||||
442 CACACTGACCTGCATGACATGACATGACATGACATGACATGACATGACATG 491
492 TCTAAGATGTTATGATGATGATGATGATGATGATGATGATGATGATGATG 541
403 nIleIleValTrpArgTyrrProThrMetSerLysLeuAlaThrLeuThrG 420
|||||:::|||||:::|||||:::|||||:::|||||
542 GATATATGTTGGAGATACCTTACGATGTCAAGTTGGCTACTCTTACGG 591
420 LysHisThrTyrgValIleuTyrrLeuAlaIleSerProAspGlyGlnThr 436
592 GACACACATFACAGAGTCTTATCTTCCATCTCACCTGATGACAGACC 641
437 IleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnValPhePr 453
|||||:::|||||:::|||||:::|||||:::|||||
642 ATACTCTGAGCTGAGATGAGACCGTAAAGTTTGGAACTGTGCC 691
453 oSerProLysSerGlnAsnThrGlnSerGluIleGlyAlaLeuSerLeuG 470
|||||:::|||||:::|||||:::|||||
692 ATCCCTTAATCTCAGAACACTGACACTGAATAGAGAGCATCTTTGG 741
470 LysArg 471
|||||
742 GAAGA 746

seq_name: gb_est1:AM030735

seq_documentation_block:

LOCUS AM030735 710 bp mRNA linear EST 18-MAY-2001

DEFINITION EST273390 tomato callus, TAMU Lycopersicon esculentum cDNA clone

ACCESSION CLEC25N23, mRNA sequence.

VERSION AM030735

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Asteridae: euasterids I: Solanales; Solanaceae; Solanum; Lycopersicon. 1 (bases 1 to 710)

AUTHORS Alcalá, J., Vredalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ann, S., Rönning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.

TITLE Generation of ESTs from tomato callus tissue

JOURNAL Unpublished (1999)

COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

1..710
location/Qualifiers

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEC25N23"

```

/clone.lib="tomato callus, TAMU"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/notes="vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Giovanni laboratory; cDNA - cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST library"
BASE COUNT      209 a      142 c      149 g      209 t
ORIGIN

alignment_scores:
  Quality: 1001.00      Length: 193
  Ratio: 5.268          Gaps: 0
  Percent Similarity: 98.446      Percent Identity: 94.819

alignment_block:
US-09-701-572-2 x AM030735 ..

Align seg 1/1 to: AM030735 from: 1 to: 710

283 AspheValSerLysLeuSerGlyHisLysSerGluValCysGlyLeu 299
|||||
3 GATTTTCAGTACGAGTGTGCTCAATCAATCAAGGTTTGGGCTCAA 52
299 STPSETYRASPAsnArgGluLeuAlaSerGlyLysAsnAspAsnLys 316
|||||
53 ATGGCTTATGATACCGTGAATTAAGTTCAGGTGGAATGATATCGCG 102
316 euepValTPAsnGlnHisSerThrGlnProValLeuLysTyrCysGlu 332
|||||
103 TTTTGTATGGAACAACATTCACACACACCTGCTGAAATCTGTGAG 152
333 HisThrAlaValAlaValAlaLeuAlaTPSerProHisLeuHisGly 349
|||||
153 CATACCTCTCGGTTAAGCCATTGCATGCTCCCCCATCTCATGAGCT 202
349 uLeuAlaSerGlyGlyThrAlaAspArgCysIleArgPheTPAsn 366
|||||
203 TCTAGCATCCGGTGGAGGACACATGTCATGATTAATTTGTGAGAC 252
366 hrThrTPAsnSerHisLeuSerCysMetAspThrGlySerGlnVal 382
|||||
253 CCACCACTAATACACATCTCATGATGACACTGCGAGTCTGTCTGT 302
383 AsnLeuValTPSerLysAsnValAlaGluLeuValSerThrHisGly 399
|||||
303 AACTCTGTGTCGAGGAATGTCATGATTAATGTCACACATGCTGTA 352
399 rSerGlnAsnGlnIleLeuValTPArgTyrProThrMetSerLysLeu 416
|||||
353 CTCTCAAAATACAGATTAATGTTGAGATATCCGACAAATGTTCAAT 402
416 lathLeuTPArgHisLysThrArgValLeuTyrLeuAlaIleSerPro 432
|||||
403 CTACTCTGAGGTCATACATTAATGAGCTTATATCTCTAATCTCA 452
433 AspGlyGlnThrIleValThrGlyAlaGlyAspGlyThrLeuArgPhe 449
|||||
453 GATGGACAGACATTTGCACTGAGACAGAGAAAGAACATTCGATTCG 502
449 pAsnValPheProSerProLysSerGlnAsnThrGlnSerGluIleGly 466
|||||
503 GAATGTTTCCCTTCTCCAAAGTCAAAAGAACCCGAGACTGAATGAG 552
466 lathSerLeuGlyArgThrThrIleArg 475
|||||
553 CATCTTCGTTGGTACAGACTCAGATCAGG 581
seq_name: gb_est2:B1456001

```

```

seq_documentation_block:
LOCUS      B1456001      872 bp      mRNA      linear      EST 21-ANG-2001
DEFINITION 603170445F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5250130 5',
mRNA sequence.
ACCESSION  B1456001
VERSION    B1456001.1 GI:15246657
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.

REFERENCE  1 (bases 1 to 872)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-remail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L14M1631 row: 1 column: 11
            High quality sequence stop: 834.
            Location/Qualifiers
                1..872
                /organism="Mus musculus"
                /strain="C57/B6"
                /db_xref="taxon:10090"
                /clone="IMAGE:5250130"
                /clone_lib="NCI_CGAP_Mam5"
                /tissue_type="tumor, gross tissue"
                /dev_stage="7 months"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Robin Humphreys,
                NIH"

BASE COUNT      174 a      256 c      268 g      174 t
ORIGIN

alignment_scores:
  Quality: 976.00      Length: 288
  Ratio: 3.920          Gaps: 6
  Percent Similarity: 86.458      Percent Identity: 68.750

alignment_block:
US-09-701-572-2 x B1456001 ..

Align seg 1/1 to: B1456001 from: 1 to: 872

157 TyrIysValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsn 173
|||||
2 TTCAAGGTGCTGAGCGCCGAGAGCTTCAGACGACTCTACCTCACTT 51
173 uValAspTPSerSerHisAsnValLeuAlaValGlyLeuLysGlnCys 190
|||||
52 GGTGACCTGTCTCTCCCAACGTCCTAGCGTGGGCTGGGCACTGGCG 101
190 aTyrLeuTPAsnAlaCysSerSerLysValThrLysLeuCysAspLeu 206
|||||
102 TGTACCTGTGAGTGCATGACACGACCGACGTCGCTGTGAGACTC 151
207 GlyValAsp...AspCysValCysSerValGlyTPAlaGlnArgGly 222
|||||
152 TCTGTGAAGGAGGACTCAGTCACTTCGCTGCTGAGCGCGGAGAA 201
222 rHisLeuAlaValGlyThrAsnAsnGlyLysValGlnIleTPAspAla 239

```

seq_name: gb_est2.BG760896

LOCUS BG760896 833 bp mRNA linear EST_15-MAY-2001
DEFINITION 602771708.F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4840604 5',
mRNA sequence.
ACCESSION BG760896
VERSION BG760896.1 GI:14071549
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 833)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-retail.nih.gov
 Tissue Procurement: ATCC/DICD/DPF
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L16C1672 row: a column: 21
 High quality sequence stop: 765.
 Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4840604"
/clone_1id="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/notice="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCGACGAG(C). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```

alignment_scores:	Quality: 917.00	Length: 2773
	Ratio: 3.869	Gaps: 6
Percent Similarity:	86.813	Percent Identity: 67.399
alignment_block:		
US-09-701-572-2 x BG508096	..	

Align seg 1/1 to: BG760896 from: 1 to: 833

[illegible]

```

311 GlysAspAsnLysLeuPheValTrpAsnGlnHisSerThrGlnProVa 327
|||||
396 GCGAACGACAAACAGCTGCTGCTGCAATCACTCAAGCCCTAGCCCT 445
327 IleuSTyrCysGlnHisThrAlaValLysAlaIleAlaTrpSerP 344
|||||
446 GCGACGTACACGACGACCTGCGGCTGGAAGCCATCGCTGCTGCC 495
344 roHisLeuHisGlyLeuLeuAlaSer.GlyGlyGlyThrAlaAspArg 360
|||||
496 CACATGACGACGAGGCTGCTGCGCTGCGGCTGGGACAGCTGACCGCT 545
360 sIleArgPheTrpAsnThrThrAsnSerHisLeuSerCysMetAsp 377
|||||
546 TATCCCTCTTCTGGACACGCTGACAGACACACTGATGATGACAG 595
377 hrgIserGlnValLysAsnLeuVal.TrpSerLysAsnValAsnGlu 393
|||||
596 GCGGCTGCCAAGTGTCAATCTGGCTGCTCCAGCAGCCACGACGCT 645
393 uAlaSerThrHisGlyTyrSerGlnAsnGlnIleValTrpArgTyrP 410
|||||
646 GGTGACGACGACGAGTACTCAGACAGACAGATCCTGTCTGGAAGTAC 695
410 roThrMetSerLysLeuAlaThrLeuThrGlyHisThrTyrArgVal 426
|||||
696 CCTCCCTGACACGAGTGGCCAGCTGACCGGACATCCATCCGCGCTG 745
427 TyrLeuAlaIleSerProAspGlyGlnThrIleValThrGlyAlaGly 443
|||||
746 TACCTGCGCAATGTC.CCTGATGCGGAGGCGCATGCTGATGCTGACA 794
443 pgluThrLeuArg 447
|||||
795 CGAGACCCCTGAGG 807

```

seq_name: gb_est2.B1652526

seq_documentation_block:

LOCUS B1652526

DEFINITION 603299431P1 NIH_CGAP_Mam3 Mus musculus cDNA IMAGE:5340103 5',

mRNA sequence.

ACCESSION B1652526

VERSION B1652526.1 GI:15566762

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 786)

AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cga@gsf-research.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LAM11863 row: n column: 08

High quality sequence stop: 783.

FEATURES

source

location/Qualifiers

1..786

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone_image="5340103"

/clone_image="NIH_CGAP_Mam3"

/tissue_type="tumor, gross tissue"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: NotI;

BASE COUNT 170 a 231 c 233 g 152 t

ORIGIN

alignment_scores:

Quality: 902.50

Ratio: 4.102

Percent Similarity: 87.649

Percent Identity: 67.331

Gaps: 251

Gaps: 3

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Gaps: 331

Site_2: Salt: Cloned unidirectionally. Primer: Oligo dr.

Average insert 2 kb. Library constructed by Life

Technologies, catalog #12017-018. Investigators providing

samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference

for transgenic model: Xu et al., Nature Genetics 22, 37-43

(1999). Note: this is a NIH_CGAP Library."

alignment_block:
US-09-701-572-2 x B1652526
Align seg 1/1 to: B1652526 from: 1 to: 786

```

221 GlyThrHisLeuAlaValGlyThrAsnAsnGlyLysValGlnIleTrpAs 237
|||||
24 GGAACCTTGTCGCGAGTAGTACACACAAAGGCTTCGTCAGATCTGGGA 73
237 PALAAlaArgCysLysLysIleArgSerMetGlnGlyHisArgLeuArg 254
|||||
74 CGCTGCTGCTGGGAGAGCTGCATGCTGAGGAGGACACACACAGAG 123
254 ALGlyAlaLeuAlaTrpSerSerSerLeuLeuSerSerGlyArgAsp 270
|||||
124 TGGGGGCGCTGGCCGGAATCTGACACAGTGTGTCATCTGATGAGCGGTGAC 173
271 LysAsnIleTyrGlnArgAspIleArgThr.....GlnGluAspPh 284
|||||
174 CGCATGATCTTACACGCGGATATCGCACACACCCCTGTCAGTACAGAG.. 221
284 eValSerLysLeuSerGlyHisLysSerGlyValCysGlyLeuLysTrp 301
|||||
222 ...CGGGGCTGCGAGGGCGACCGGACAGAGTGTGGCTTAAGTGT 267
301 eTyrAspAsnArgGluLeuAlaSerGlyGlyAspAsnLysLeuPhe 317
|||||
268 CCACAGACACACAGCTGCTGCTGCGGCGGACATGACAAACCTGCTC 317
318 ValTrpAsnGlnHisSerThrGlnProValLeuLysTyrCysGlnHisTh 334
|||||
318 GTGTGGAACCACTAGTCTGAAGCCCTGTGCAGCAGATATGAGAGACACT 367
334 rAlaAlaValLysAlaIleAlaTrpSerProHisLeuHisGlyLeuAla 351
|||||
368 GGCACGCTGAAGGCTATTGCTGCTGCCACACACAGCATGAGCTGCTGG 417
351 lAserGlyGlyThrAlaAspArgCysIleArgPheTrpAsnThr 367
|||||
418 CATCTGCTGCTGCGACGCGCTGACCCCTGCATCCATTTGTGAACACTTG 467
368 ThrAsnSerHisLeuSerCysMetAspThrGlySerGlnValCysAsnLe 384
|||||
468 ACAGGTCAAGCACTGACATGATGACACAGGCTCACAAGTGTGCAACT 517
384 uValTrpSerLysAsnValAsnGluLeuValSerThrHisGlyLysSerG 401
|||||
518 GGCCTGTGTCAAAGCAGCCCAATGAGCTGTAGACACATGAGCTACTCAC 567
401 lAsnGlnIleIleValTrpArgTyrProHisMetSerLysLeuAlaThr 417
|||||
568 AGAACCAATCTCTCTGTGGAAGTACCCGCTTACCCAGAGTGGCCAG 617
418 LeuThrGlnHisThrTyrArgValLeuTyrLeuAlaIleSerProAspG 434
|||||
618 CTCACCTGCGCACTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 667
434 yGlnThrIleValThrGlyAlaGlyAspGluThrLeuArgPheTrpAsnV 451

```


KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 860)
JOURNAL NIH-MGC <http://mgs.nci.nih.gov/>.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: L1AM12219 row: p column: 13
High quality sequence stop: 668.

FEATURES
source
Location/Qualifiers
1..860
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5534076"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site_1: Not;
Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb.
281 c 280 g 138 t

BASE COUNT 161 a 281 c 280 g 138 t
ORIGIN
alignment_scores:
Quality: 895.00 Length: 281
Ratio: 3.874 Gaps: 7
Percent similarity: 82.206 Percent identity: 64.769

alignment_block:
US-09-701-572-2 x BM462011 ..

Align seg 1/1 to: BM462011 from: 1 to: 860

```

125 SerMetHisSerLeuSerProPheMetAspAspPheValProGlyVal 141
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 TCTCCCTACTCCCTCTCTCCGTCAGCAACAAGAC.....CAGAGCT 73
141 IAsnHisSerProValLysAlaProArgLysValProArgSerProTyrL 158
: :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
74 GCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 123
158 yValLeuAspAlaProAlaLeuGlnAspAspPheTyrLeuAsnLeuVal 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 AGGTCTCTGACGCGCCGACGTCGACGACCTCTTACCTCAATCTGGTG 173
175 AspTrpSerSerHisAsnValLeuAlaValGlyLeuGlnAsnValTyr 191
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 GACTGCTGCTCCCTCAATGTCGTCAGCTGGGCTGAGGCACCTGCTGTA 223
191 rLeuTrpAsnAlaCysSerSerLysValThrLysLeuCysAspLeuGly 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 CCTGGAGATGCTGTCACGACGACGACGCTGTCGTCGTCGTCGTCGTA 273
208 aAsp...AspCysValCysSerValGlyTrpAlaGlnArgGlyThrHis 223
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 TGGAGAGGAGCTCAGTGACCTCGTGCGTGGTCTGACCGGGGAACTGG 323
224 LeuAlaValGlyThrAsnAsnGlyLysValGlnIleTrpAspAlaAla 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 GTGGCGGTGGGACACACAGAGGCTTGTGTCGATCTGGAGCGAGCGCG 373

```

```

240 gCysIlySIIeArgSerMetClnGlnHisArgLeuArgValGlyAla 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 AGGAGAGAGAGCTGTCCATGTGTGAGGCGCACAGCGACGCTGGGGCGC 423
257 euAlaTrpSerSerLeuLeuSerSerGlyLysArgSpsIysnle 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 TGGCCGGAATGCTGACGAGCTGCTCGGAGCGCGGACCGCAGTGAAC 473
274 TyrGlnArgAspIleArgThr.....GlnGluAspPheValSerGly 287
474 CTGCAGAGGAGACATCCGACACCCCGCACGTCAGTGGAG.....CGGG 517
287 sLeuSerGlnHisLysSerGlnValCysGlyLeuLysTrpSerTyrAsp 304
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
518 GTCGACGGGCGCACCGGACAGAGTGTGCGGCTCAAGTGTCCACAGACC 567
304 snArgGlnLeuAlaSerGlyLysAsnAspAsnLysLeuPheValTrpAs 320
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
568 ACCAGCTCTCGCTCGGGGGGCGACACACAGCTGCTGCTGGAT 617
321 GlnHisSerThrGlnProValLeuLysTyrCysGlnHisThrAlaAla 337
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
618 CACTCGAGCGCTGAGCCCGTCGACAGTACAGCGACACCTGGCGCGCT 667
337 lLysAlaIleAlaTrpSerProHisLeuHisGlyLeuAlaAlaSer 353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
668 GAAGGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 717
354 GlnGlyThrAlaAspArgCysIleArgPheTrpAsnThrThrAsnSe 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
718 GGGGGCACACTACCGCTGATCCGCTTCTGTAACACCTGACAGAGA 767
370 rHisLeuSerCysMetAspThrGlySerGlnValCysAsnLeuValTrp 386
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
768 ACCACTGACGTATGACACAGCGGCTCCCAAGTGTCAATCTGGGCTG 817
387 SerLysAsnValAsnGlnLeu..ValSerThrHisGly 398
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
818 TCCAGACGACGCCAACCAAGCTGGGTGAGCAGCGACGG 855

```

seq_name: gb_estl:A1895812

seq_documentation_block:

LOCUS A1895812 501 bp mRNA linear EST 18-MAY-2001
DEFINITION EST265255 tomato callus, TAMU Lycopersicon esculentum cDNA clone
CLC10M13, mRNA sequence.

ACCESSION A1895812
VERSION A1895812.1 GI:5601714

KEYWORDS
SOURCE EST.

ORGANISM

tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; eunasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 501)
Alcala,J., Vredalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.

AUTHORS Ilang,F., Upton,J., Craven,M.B., Bowman,C.L., Ann,S., Koning,
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)

TITLE

CONTACT: CUGI

COMMENT Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source

1..501
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLC10M13"

```

/clone.lib="tomato callus, TAMU"
/tissue.type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      145 a      102 c      113 g      141 t
ORIGIN

alignment_scores:
  Quality: 892.00      Length: 166
  Ratio: 5.373      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 97.590

alignment_block:
US-09-701-572-2 x AI895812 ..

Align seg 1/1 to: AI895812 from: 1 to: 501

283 AspPheValSerLysLeuSerGlnHisLysSerGluValCysGlyLeu 299
|||||
3  GAATTTGTCAGTAAGCTGAGTGTCAATTAATCAGAGGTTTGCGCTCAA 52

299 STPSPSTYrAspAsnArgGluLeuAlaSerGlyGlyAsnAspAsnLys 316
|||||
53 ATGGCTTATGATTAACCGGTAATTAAGCTTCAGCTGGAATGATAACGCGC 102

316 eupheValTPraSngInHisSerThrGlnProValLeuLysTYrCysGlu 332
|||||
103 TTTTGTATGAGAACACACATCAACACACACTGCTGAAATACCTGTGAG 152

333 HisThrAlaAlaValLysAlaIleAlaTrpSerProHisLeuHisGly 349
|||||
153 CATACGCTGCGCGGTAAAGCCATTCATGTCATGCCCCCATCTCATGGCT 202

349 uLeuAlaSerGlyGlyThrAlaAspArgCysIleArgPheTrpAsnT 366
|||||
203 TCTAGCATCCGGTGGAGGACACGCTGATCATGATTAATTCCTGGAGAA 252

366 hThrThrAsnSerHisLeuSerCysMetAspThrGlySerGlnValCys 382
|||||
253 CCACACATATACACATCTCAGTTGCATGGACACCTGGCAGTCAAGTCT 302

383 AsnLeuValTrpSerLysAsnValAsnGluLeuValSerThrHisGly 399
|||||
303 AATCTGTGTGTGCGAAGATGTCATTAATTAAGTCAGCACACATGGTTA 352

399 rSerGlnAsnGlnIleIleValTrpArgTrpProHisMetSerLysLeu 416
|||||
353 CTCCTCAAAATCAGATATAGTTGGAGATATCCGACAAATGCTTAAGATG 402

416 lathrLeuThrGlyHisThrTYrArgValLeuTYrLeuAlaIleSerPro 432
|||||
403 CTACTCTGACAGGTACATATAGAGTCTTAATCTTGTATATCTCA 452

433 AspGlyGlnThrIleValThrGlyAlaGlyAspGluThrLeuArgPhe 448
|||||
453 GATGGACAGCAATGTCCTGAGCAGAGAGATGAACACTTCGATTC 500

seq_name: gb_est2:BI554814

seq_documentation_block:
LOCUS      BI554814      863 bp      mRNA      linear      EST 05-SEP-2001
DEFINITION  603236792F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5289670 5',
mRNA sequence.
ACCESSION  BI554814
VERSION    BI554814.1  GI:15442128
KEYWORDS   EST.

```

```

SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   1 (bases 1 to 863)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cga@pds-remail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1732 row: h column: 23
            High quality sequence stop: 855.
FEATURES
  source
    1..863
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="IMAGE:5289670"
    /clone.lib="NIH_CGAP_Mam3"
    /tissue.type="tumor, gross tissue"
    /lab_host="DH10B"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
    Average insert 2 kb. Library constructed by Life
    Technologies, catalog #12017-018. Investigators providing
    samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
    (1999). Note: this is a NCI_CGAP Library."
BASE COUNT      191 a      236 c      257 g      159 t
ORIGIN

alignment_scores:
  Quality: 876.50      Length: 249
  Ratio: 4.039      Gaps: 5
  Percent Similarity: 87.149      Percent Identity: 67.871

alignment_block:
US-09-701-572-2 x BI554814 ..

Align seg 1/1 to: BI554814 from: 1 to: 863

226 ValGlyThrAsnAsnGlyLysValGlnIleTrpAspAlaAlaArgCys 242
|||||
3  GTAGTACACACAAGAGGCTTGTGTCAGATGCTGGAGCCCTGCTGGGA 52

242 slyIleArgSerMetGluGlyHisArgLeuArgValGlyAlaLeuAla 259
|||||
53 GAACTGTCCATGCTGGAGGGCCACACAGCAGCGATGGGGCCCTGGCCT 102

259 rPserSerSerLeuLeuSerSerGlyGlyLysArgAspLysAsnIleTYrGln 275
|||||
103 GGATGCTGACCAATTTCTCATCTGTGAGCCGTGACCCGACATCTCAG 152

276 ArgAspIleArgThr.....GlnGluAspPheValSerLysLeu 289
|||||
153 CGGATATCCGACACACACCCCTGACAGTACAG.....CGGCGCTGCA 196

289 rGlyHisLysSerGluValCysGlyLeuLysTrpSerTYrAspAsnArg 306
|||||
197 GGGCCACCGCAGGAGAGTGTGTGCTTAAGTGTCCACAGCACCACAG 246

306 lueuAlaSerGlyGlyAsnAspAsnLysLeuPheValTrpAsnGlnHis 322
|||||
247 TGCTTGCTGCTGGGGGCAATGACAAAGCTGCTGATGAGAACCACTCT 296

323 SerThrGlnProValLeuLysTYrCysGlnHisThrAlaAlaValLys 339

```

```

297 AGCTAAGCCCTGTGACAGATATACGAGACACCTGGACCGGTGAAGC 346
339 ATLEATPSPERPROHISLEUHSISGLYLEUHALEUASERGLYGLYT 356
347 TATTGGCTGGTCCACACACAGATGACTGCGCATGTGGGGGGA 396
356 hvalaasparcysllearghetpaspntrhrthrasnsershisleu 372
397 CGGCTGACCGCTGCATCTGATGACACCTGACAGCTGACAGCCACG 446
373 SerCysMetAspThrGlySerGlnValCysAsnLeuValTrpSerLys 389
447 CAGTGATTTGACACAGGCTCACAAGTGTACACCTGGCTGGCCAGA 496
389 nvalaasngluLeuValSerThrHisGlyTrpSerGlnAsnGlnIle 406
497 CGGCATGAGCTGTGAGACACATGCTACTACAGAACACAGATCTCG 546
406 altrpargtyrProthMetSerLysLeuAlaThrLeuThrGlyHis 422
547 TGTGAAGTACCGCTCCCTTACGACAGTGGCCAGCTCACTGGCCAC 596
423 TyrArgValLeuThrLeuAlaIleSer.ProaspGlyGlnThrIle 439
597 TATGCTGCTCTACTGCTGACATGCTCCCTGATGGGAGGCCATGTA 646
439 hrcglvalaaglyaspGluThrLeuArgPheTrpAsnValPhe.Pro 455
647 CGGAGCTGAGATGAGACCTGAGCTTCTGATGCTCTTCAGGCAAC 696
455 OlysserGlnAsnThrGluSerGluIleGlyAlaLeuSerLeu 697
697 ACCTCT.....ACAAGGAATCTGTGTGTGTCAACCTC 733

```

seq_name: gp_est2:BE790506

seq_documentation_block:

LOCUS BE790506 888 bp mRNA linear EST 20-SEP-2000
 DEFINITION 60158211P1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3936770 5',
 mRNA sequence.

ACCESSION BE790506 GI:10211704

VERSION BE790506.1 GI:10211704
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 888)
 NIH-MGC http://mgi.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgsb@nsi.nih.gov

Tissue Procurement: DCTD/BTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov

Plate: LNCM/81 row: F column: 03

High quality sequence stop: 777.

Location/Qualifiers

1..888

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone IMAGE:3936770"

/clone_lib="NIH_MGC_7"

/tissue_type="small cell carcinoma"

/cell_line="MGC3"

/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 166 a 276 c 307 g 139 t
 ORIGIN

alignment_scores:

Quality: 868.00 Length: 270

Ratio: 3.910 Gaps: 9

Percent Similarity: 82.222 Percent Identity: 65.185

alignment_block:

US-09-701-572-2 x BE790506 ..

Align seg 1/1 to: BE790506 from: 1 to: 888

150 ArgLysValProArgSerProTyrLysValLeuAspAlaLeuG 166

13 CGCAAGATCTCCAAAGATCCCTCAAGGTGCTGGACGCCGAGCTGCA 62

166 nAspAspPheTyrLeuAsnLeuValAspTrpSerSerHisAsnValLeu 183

63 GGACGACTTCTACCTCAATCTGTGTGACTGCTGCTCCTCAATGTGCTCA 112

183 LavalGlyLeuGlyAsnValTyrLeuTrpAsnAlaCysSerLys 199

113 GCGTGGGCTAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 162

200 ValThrLysLeuCysAspLeuGlyValAsp...AspCysValCysSerVa 215

163 GTGACGCGGCTCTGTGACCTCTCACTGAGAGGAGGAGGAGGAGGAGCT 212

215 LglYTrpAlaGlnArgGlyThrHisLeuAlaValGlyThrAsnGln 232

213 GGGCTGTGTGTGAGCGGGGAGACCTGTGTGTGTGTGTGTGTGTGTGTGT 262

222 ySvalGlnIleTrpAspAlaIleArgCysLysIleArgSerMetGlu 248

263 TGTGACAGATCTGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCT 312

249 GlnHisArgLeuArgValGlyAlaLeuAlaTrpSerSerLeuLeu 265

313 GGCCACAGGAGGACCGGCTGGGCGCTGGCTGGAATGCTGAGACGCTGC 362

265 rSerGlyGlyArgAspLysAsnIleTyrGlnArgAspIleArgThr... 280

363 GTCCGGGAGCGCGACCCATGATCTCTGACAGAGGACATCCGACCCCGC 412

281GlnGluAspPheValSerLysLeuSerGlnHisLysSerGluVal 295

413 CACTGCAGCTGGAG.....CGGCGGCTGCAAGGCCACCGGACGAGGCTG 456

296 CysGlyLeuLysTyrSerTyrAspAsnArgGluLeuAlaSerGlyGlyAs 312

457 TCGGGGCTCAAGTGTGTCACAGACACACAGACGCTCGGCTGGGGGGA 506

312 nAspAsnLysLeuPheValTrpAsnGlnHisSerThrGlnProValLeu 329

507 CGACAAACAGCTGCTGTGTGAAATCACTGAGCCCTGAGCCCTGAGCG 556

329 ySTyrCysGlnHisThrAlaValLysAlaIleAlaTrpSerProHis 345

557 AGTACAGGAGACCTGGCGGCGTGAAGGCCATGCTGTGCTCCCATAT 606

346 LeuHisGlyLeuLeuAlaSerGlyGlyGlyThrAlaAspArgCys...Ile 362

607 CAGACAGGCGCTGCTGCTGCGGGGGGCGGACAGCTGACCGCTGTAATCC 656

362 rPheTrpAsnThrThrAsnSerHisLeuSerCysMetAspThr...G 378

```

|||||
657 GCTTCGGAACACGCTGACAGACACACCTGACGGTTATGACACGGG 706
|||||
378 lYserGlnValCysAsnLeuValTTPSerLysAsnValAsnGluLeuVal 394
|||||
707 GCTCCCAATGTGCAATCTGGCTGTCCACACGACCAACAGCTGGTG 756
|||||
395 SerThrHisGly.....TyrSerGlnAsnGlnIleValTTPArgTy 409
|||||
757 AAG...CACGGCAGGGTAACCTCCACAGACAGAGA...TCTTGTGGAGAGTC 800
|||||
409 rPro 410
|||||
801 CCT 804

seq_name: gb_est2:BI153551

seq_documentation_block:
LOCUS BI153551 873 bp mRNA linear EST 05-JUL-2001
DEFINITION 602917131F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5067822 5',
mRNA sequence.
ACCESSION BI153551
VERSION BI153551.1 GI:14613552
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 873)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L14M1182 row: m column: 07
High quality sequence start: 15
Location/Qualifiers
1. 873
/organism="Mus musculus"
/strain="C2EHC II (fetal)"
/db_xref="taxon:10090"
/clone="IMAGE:5067822"
/clone_id="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dm.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 186 a 263 c 263 g 162 t
ORIGIN

alignment_scores:
Quality: 866.00 Length: 240
Ratio: 4.144 Gaps: 2
Percent Similarity: 87.083 Percent Identity: 68.333

alignment_block:
US-09-701-572-2 x BI153551 ..

Align seg 1/1 to: BI153551 from: 1 to: 873

231 GlyValGlnIleTTPAspAlaAlaArgCysLysLysIleArgSerMet 247

```

```

|||||
29 GCTTCGTCGACGATCTGGAGCGCTGCTGGGAGAGAGCTGTCATGCT 78
|||||
247 tGluGlyHisArgLeuArgValGlyAlaLeuAlaTTPSerSerLeuL 264
|||||
79 GGAGGGCCACACACACAGAGTGGGGGCGCTGGCTGGAAATGACAGT 128
|||||
264 euseSerGlyLArgAspLysAsnIleTyrGlnArgAspIleArg..T 280
|||||
129 TGTCATCTGTAGCCGTGACCGCATGATCTTAACGGGGATATCCGACA 178
|||||
280 hrgIngluAspPheValSerLysLeuSerGlyHisLysSerGluVal 296
|||||
179 CCACCCCTGCGAGTCAGCGGGCGCTGCAGGGCCACCGCAGAGTGTCT 228
|||||
297 GlyLeuLysTTPSerTyrAspAsnArgGluLeuAlaSerGlyLys 313
|||||
229 GGCTTAAGTGTCCACACACACACCTGCTGCTCGGGGGCAATGA 278
|||||
313 PAsnLysLeuPheValTTPAsnGlnHisSerThrGlnProValLeuLys 330
|||||
279 CAACAAGCTGCTGCTGTGAAACCACTACTAGTCAAGCCCTGTGACAGT 328
|||||
330 yTCysGluHisThrAlaAlaValLysAlaIleAlaTTPSerProHis 346
|||||
329 ATACGAGACACCTGCGAGCGGTGAAGGCTATGCTGCTGCCACACAG 378
|||||
347 HisGlyLeuLeuAlaSerGlyGlyGlyThrAlaAspArgCysIleArg 363
|||||
379 CATGAGCTGCTGGCATGTGGTGGGCGCGGCTGACCGCTGCATTCGATT 428
|||||
363 eTTPAsnThrThrAsnSerHisLeuSerCysMetAspThrGlySerG 380
|||||
429 CTGGAACACTGTGACAGCTCAGCAGCATGTCGATGACACAGGCTCAC 478
|||||
380 lValCysAsnLeuValTTPSerLysAsnValAsnGluLeuValSerThr 396
|||||
479 AAGGTCAACACCTGGCTGTCCAAAGCAGCCCAATGAGCTGGAGACACA 528
|||||
397 HisGlyTyrSerGlnAsnGlnIleIleValTTPArgTyrProThrMet 413
|||||
529 CATGGCTACTACAGAACCAATCTGCTGTGAAGTACCCGCTTAC 578
|||||
413 lYsLeuAlaThrLeuThrGlyHisThrTyrArgValLeuTyrLeuAla 430
|||||
579 GCAGTGGCCAGAGCTCAGTGCACATCATGTCGTCCTTACTCTGACCA 628
|||||
430 lSerProAspGlyGlnThrIleValThrGlyAlaGlyAspGluThrLeu 446
|||||
629 TGTCCTCTGATGGGAGGCGCATAGTCAACCGAGCTGAGATGAGACCTG 678
|||||
447 ArgPheThrAsnValPheProSerProLysSerGlnAsnThrLysSerG 463
|||||
679 AGGTTCTGGAATGTCTTC.....AGCAAAACACGCTCTCAAGAAATC 722
|||||
463 uileGlyAlaLeuSerLeu 469
|||||
723 TGTCCTGTGCTCAAGCTC 741

seq_name: gb_est2:BI904107

seq_documentation_block:
LOCUS BI904107 756 bp mRNA linear EST 16-OCT-2001
DEFINITION 603166733F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5254579 5',
mRNA sequence.
ACCESSION BI904107
VERSION BI904107.1 GI:16166293
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 756)
REFERENCE

```

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LMA11643 row: b column: 20
 High quality sequence start: 68
 High quality sequence stop: 756.
 Location/Qualifiers

FEATURES

source

1. 756
 /organism="Mus musculus"
 /strain="CZECH II"
 /db_xref="taxon:10090"
 /clone="IMAGE:5254579"
 /clone_lib="NCI CGAP Lu33"
 /tissue_type="Pooled lung tumors"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: Lung; Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dT) primer [5', TGTACCAACTGAGAGGAGGAGCGCGCCCTCTTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3D vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 216 c 237 g 152 t 2 others
ORIGIN

alignment_scores:
 Quality: 862.00 Length: 244
 Ratio: 4.124 Gaps: 3
 Percent Similarity: 85.656 Percent Identity: 69.262

alignment_block:

US-09-701-572-2 x B1904107 ..

Align seg 1/1 to: B1904107 from: 1 to: 756

```

159 ValIeuAspAlaProAlaIeuGlnAspAspPheTyrLeuAsnLeuValAs 175
|||||
32 GTGCTGGACGCCGCCAGAGCTTCTTA CGATTCTACCTCAACTGGTGA 80
175 pTTPSerHisAsnValIeuAlaValAlaGlyLeuGlyAsnValTyrL 192
|||||
81 CTGGCTCTCCCTCAACGCTCTCAGTGTGGGCTGGGACACCTGGTGTACC 130
192 eUTTPAsnAlaCysSerSerLyValThrLysLeuGlyAsnLeuGlyVal 208
|||||
131 TGTGGAGTGCATGACACAGCCAGGTGACCCGCTCTGTACCTCTCTGTA 180
209 Asp...AspCysValCysSerValGlyTPAlaGlnArgGlyThrHisLe 224
|||
181 GAAGGGGACTCACTGACTTCCTGCTGTGCTGTGACGGGGGAACTTGST 230
224 uAlaValGlyThrAsnAsnGlyLysValGlnIleTPAspAlaAlaArgC 241
|||||
231 CGCAGTAGGTACACACAGGCGCTGTCGACAGATCTGGGACGCTGCTG 280
241 yAlaLysLysIleArgSerMetGluGlyHisArgLeuArgValAlaLeu 257
|||||
281 GGAAGAGAGCTTCATGCTGTGAGGGGCCACACAGACAGAGTGGGGCGCTG 330

```

```

258 AlaTPSerSerSerLeuLeuSerSerGlyGlyArgAspLysAsnIleTyr 274
|||||
331 GCCTGGAATGCTGACACAGATTGATCTGTGATCCCTGACCCGATGATCCT 380
274 rGlnArgAspIleArgThr.....GlnLysPheValSerLysL 288
|||||
381 ACAGCGGATATCCGACACACACCCCTGCATGACAG.....CGCGGC 424
288 euSerGlyHisLysSerGluValCysGlyLeuLysTyrSerTyrAspAsn 304
|||||
425 TGCAGGGCCACCGGACAGAGAGTGTGGCTTAAAGTGTCCACAGACAC 474
305 ArgGluLeuAlaSerGlyGlyAsnAspAsnLysLeuPheValTPAsnG1 321
|||
475 CAGTGTCTTGCCTCGGGGGCAATGACAAACAGCTGCTGTGTGAACCA 524
321 nHisSerThrGlnProValLeuLysTyrCysGlnHisThrAlaVal 338
|||
525 CTCTAGTCTAAGCCTGTGCACAGATATACGAGCACCCTGGAGCGGTGA 574
338 yAlaIleAlaTPSerProHisLeuHisGlyLeuLeuAlaSerGlyGly 354
|||||
575 AGCTATTGCTGCTGCTCCACACACAGCATGAGTCTGTGATCTGTGTGT 624
355 GlyThrAlaAspArgCysIleArgPheTPAsnThrThrAsnSerHis 371
|||||
625 GGCACGGCTGACCGCTGCATCGATTCTGAAACACTGTGACAGGTGACGC 674
371 sLeuSerGlyMetAspThrGlySerGlyValCysAsnLeuValTPSerL 388
|||||
675 ACTGCAGTGCATTTGACACAGGCTCAAGTGTGCAACCTGGCGCTGTCCA 724
388 yAsnValAsnGluLeuValSerThrHisGly 398
|||||
725 AGCAGCCCAATGAGCTGTGAGACACATGGC 756

```

